

AD-A139 268

EXPERIMENTAL DESIGN: REVIEW AND COMMENT(U) WISCONSIN
UNIV-MADISON MATHEMATICS RESEARCH CENTER
D M STEINBERG ET AL. FEB 84 MRC-TSR-2639

1 / 1

UNCLASSIFIED

DAAG29-80-C-0041

F/G 12/1

NL

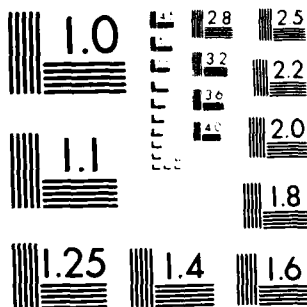
END

DATE

FILED

4 -84

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A139268

MRC Technical Summary Report #2639

EXPERIMENTAL DESIGN:
REVIEW AND COMMENT

David M. Steinberg
and
William G. Hunter

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

February 1984

(Received December 23, 1983)

DTIC
ELECTE
MAR 22 1984
S D
B

Approved for public release
Distribution unlimited

DTIC FILE COPY

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

84 11 22 101

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

EXPERIMENTAL DESIGN: REVIEW AND COMMENT

David M. Steinberg and William G. Hunter

Technical Summary Report #2639

February 1984

ABSTRACT

We review major developments in the design of experiments, offer our thoughts on important directions for the future, and make specific recommendations for experimenters and statisticians who are students and teachers of experimental design, practitioners of experimental design, and researchers jointly exploring new frontiers. Specific topics covered are optimal design, computer-aided design, robust design, response surface design, mixture design, factorial design, block design, and designs for nonlinear models.

AMS (MOS) Subject Classifications: 62K, 62K05, 62K10, 62K15

Key Words: Experimental Design, Optimal Design, Computer-Aided Design,

Robustness, Response Surfaces, Mixtures, Factorial Design,

Blocking, Nonlinear Models, Consulting, Teaching

Work Unit Number 4 - Statistics and Probability

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. The research of David Steinberg was based upon work supported by the National Science Foundation under Grant No. MCS-8210950.

SIGNIFICANCE AND EXPLANATION

Statistics is concerned not just with the analysis of data, but also with how data are collected. Such concern is quite natural, for analysis cannot be truly successful unless the data are informative, and the best way to assure informative data is to apply statistical principles to the way in which they are collected. One area in which statistics has made notable contributions to collecting informative data is experimental design. This paper provides a general, non-technical review of statistical work in experimental design, discusses important directions for future research, and makes some specific recommendations.

The focus of the paper is on those areas of experimental design that are most useful in the physical, chemical, and engineering sciences. Specific topics covered are:

1. Optimal design, in which specific criteria are developed and studied in an attempt to derive designs that will provide the experimenter with the most precise inferences possible.
2. Computer-aided design, in which computer algorithms that find optimal designs, or designs with other desired properties, are developed.
3. Robust design, which studies the sensitivity of experimental designs to departures from assumptions on which those designs were based.
4. Response surface designs, which exploit a simple, sequential experimental strategy to explore the relationship between a response variable and several continuous inputs.
5. Mixture designs, which provide strategies for modeling a response that depends on the relative amounts of several continuous inputs.
6. Factorial design, which emphasizes the usefulness of varying several factors simultaneously in an experiment, rather than just one factor at a time, and proposes economical strategies to do so.
7. Block designs, which offer efficient schemes for comparing several different treatments.
8. Designs for nonlinear models, which suggest useful ways to design experiments when the response is assumed to be a nonlinear function of unknown parameters.

The paper concludes by addressing several recommendations to experimenters and statisticians to encourage the increased use of statistically designed experiments and to facilitate the interchange of ideas between statisticians and experimenters that is an essential stimulus to future research in this area.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.



| |
|--|
| <input checked="checked" type="checkbox"/> |
| <input type="checkbox"/> |
| <input type="checkbox"/> |
| Codes |
| 3/or |
| 1 |

A-1

EXPERIMENTAL DESIGN: REVIEW AND COMMENT
David M. Steinberg and William G. Hunter

1. INTRODUCTION

Fisher's pioneering work at Rothamsted Experimental Station in the 1920's and 1930's firmly established the role of statistics in experimental design and, vice versa, the role of experimental design in statistics. His monumental work was guided by the key insight that statistical analysis of data could be informative only if the data themselves were informative, and that informative data could best be assured by applying statistical ideas to the way in which the data were collected in the first place. In the process, Fisher radically altered the role of the statistician: from one of after-the-fact technician to one of active collaborator at all stages of an investigation.

Fisher was employed to analyze data from studies conducted at Rothamsted, but he soon realized that some important questions could not be answered because of inherent weaknesses in the planning of many of the experiments. In fact, in one particularly unfortunate instance, he said that the only analysis he could perform was a post-mortem to find out why the study had died. Box (1980) described Fisher's work on the design of experiments, and how much of it was inspired by problems of field experimentation: he developed his insights concerning randomization, blocking, and replication; he invented new classes of experimental designs; he worked together with scientists who applied his ideas in their experiments; by mail, he advised experimenters in other places; and he wrote about his ideas to help investigators realize

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. The research of David Steinberg was based upon work supported by the National Science Foundation under Grant No. MCS-8210950.

richer harvests of information from their investments in experimental work. The collaborative way in which Fisher worked with the scientists at Rothamsted, aiding them in their experimental research and then using his experiences as the motivation for important statistical research, still serves as a model for statisticians to emulate. For a more detailed account of Fisher's work at that time, see Yates and Mather (1963), Yates (1964), Box (1978), and the tributes to Fisher in Biometrics (1962, Volume 18, 437-454).

Since Fisher first introduced statistical principles of experimental design, much useful statistical research has been done. Our primary purpose in this article is to provide a summary of selected work in experimental design, rather than an exhaustive review of the literature, and to offer some thoughts about future directions. Wherever possible, we will refer the reader to books that discuss the basic ideas of experimental design and present many of the most widely used plans and to other good review articles summarizing research on experimental design and highlight only the most recently published results. (See Hahn 1982b for a useful review of available books on experimental design.) We will focus especially on the design of experiments in the physical, chemical, and engineering sciences.

During the last quarter century, many papers on the design of experiments have appeared in Technometrics. Given the jubilee nature of this article, we felt it would be appropriate to begin by reading those papers, a task that proved both enjoyable and rewarding. Our summary of this research, which is presented in Section 2, provides a perspective from which to evaluate current research. In Sections 3-10 we review research in several major areas: optimal design, computer-aided design, design robustness and design sensitivity, response surface designs, mixture designs, factorial designs, block designs, and designs for nonlinear models. In Section 11 we discuss

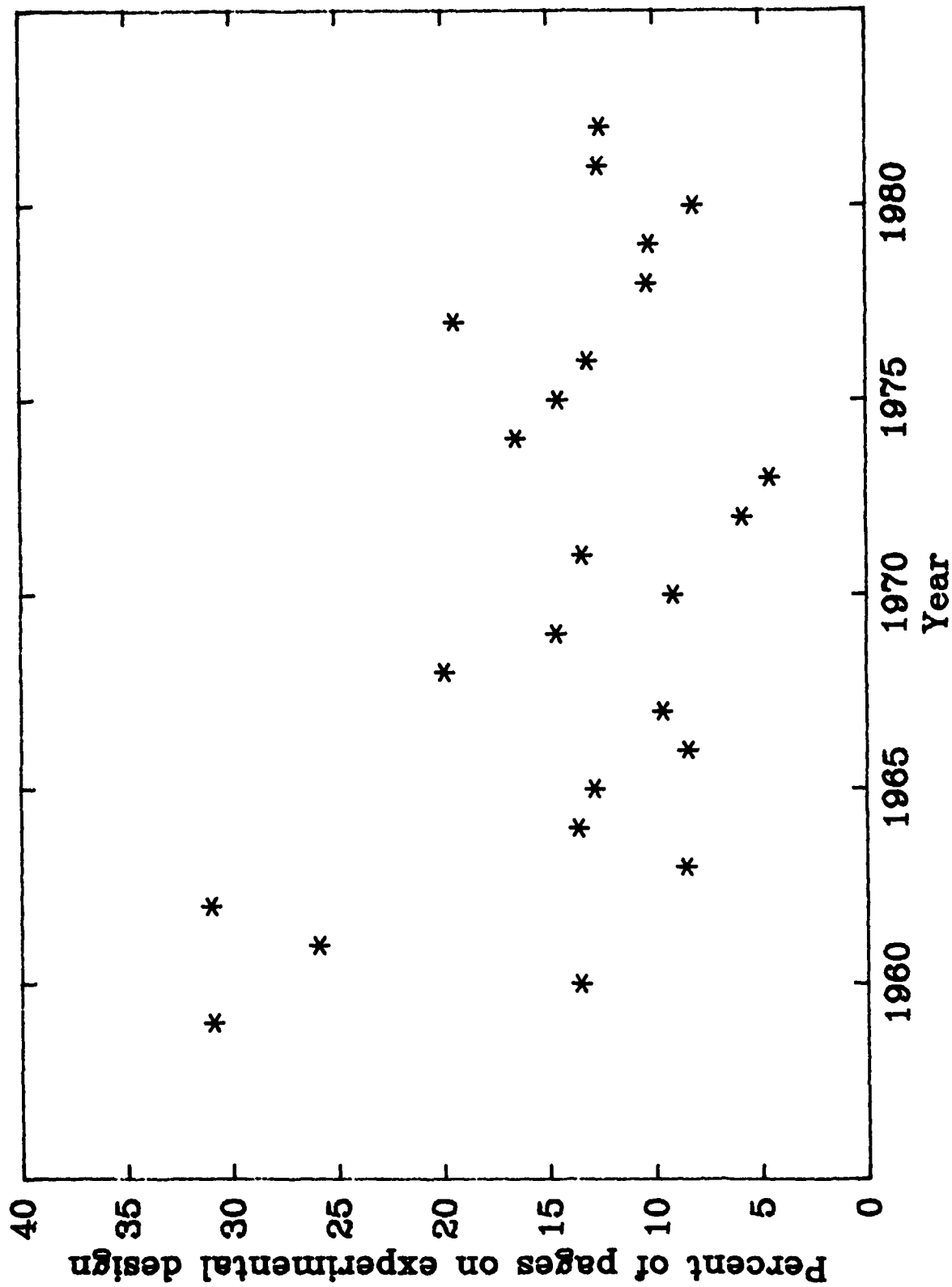
some topics that we believe deserve further study, and we offer some personal reflections on future directions. We conclude in Section 12 with some recommendations for experimenters and statisticians.

2. EXPERIMENTAL DESIGN IN TECHNOMETRICS

Experimental design has always been a prominent topic in Technometrics. Figure 1 shows the percentage of pages in Technometrics devoted to this subject on a yearly basis from 1959 through 1982. (We have included here all papers and notes whose principal focus is on the theory or technique of designing experiments, excluding papers that deal only with the analysis of particular types of designs. We have not counted book reviews, letters to the editor, corrigenda, and other editorial material.) The first years of Technometrics witnessed a profusion of articles on experimental design, which occupied 20% to 30% of the space in the journal. While such a high concentration has not been maintained in subsequent years, the percentage has still consistently exceeded 10%. By contrast, the Journal of the American Statistical Association devoted less than 1% of its pages in 1982 to articles on experimental design; the corresponding figure for the Annals of Statistics in 1982 was 4%.

Thus experimental design has clearly been a topic of major interest for Technometrics. An informative picture of the development of research on experimental design and its applications emerges from an examination of these articles. The first issues of Technometrics included many articles on factorial and fractional factorial designs and block designs, topics that were originally explored in the context of agricultural experimentation. Their appearance in Technometrics marked a realization that these concepts are also important in the physical, chemical, and engineering sciences. Moreover, the

Figure 1: Experimental Design in TECHNOMETRICS



confrontation of existing ideas in experimental design with new areas of application sparked creativity. Innovative modifications and extensions of classical experimental designs were developed and many useful articles were published in a short time. Following this initial period of enthusiasm, articles on these particular topics have continued to appear, but only sporadically.

One topic that has received continuing attention in Technometrics is the design of response surface experiments. Response surface methodology was stimulated by problems arising in chemistry and chemical engineering, in particular how to improve the performance of systems by modifying the settings of process variables (Box and Wilson 1951). The strategy advocated was to use a sequence of simple experimental designs to locate and then explore regions that promised high levels of performance. The basic building blocks were directly borrowed from or were extensions of the classical factorial designs initially used in agriculture and biology. The new conceptual framework offered by response surface methodology, especially its appeal to geometric ideas, stimulated much new research. Technometrics was a natural forum for the discussion of these new ideas. Response surface methodology provides an example of the type of stimulation that was provided by the appearance of this new journal in the statistics literature. A steady flow of articles on response surface design appeared throughout the 1960's and into the 1970's, but has abated in recent years.

Three new subjects assumed prominence in the 1970's: optimal design, computer-aided design, and mixture design. Two related topics that have come to the fore in the last 10 years are design robustness and design sensitivity. Research work in robustness and sensitivity is important because experiments sometimes must be planned in the face of a considerable degree of model uncertainty.

The emergence of optimal design as a central concern can be seen quite clearly in Technometrics: through 1970 only two articles dealt explicitly and primarily with optimal design, but since 1970 that number has increased to more than a dozen. Interest in computer-aided design has grown for two reasons: advances in computer technology and the increasing influence of optimal design theory. Much of the work in computer-aided design has gone into the development of powerful algorithms for finding optimal designs and other designs with certain desired properties (e.g., orthogonality of some factor effects, or particular confounding patterns). Mixture designs, although first discussed in the 1950's, received little attention prior to 1970; however, they have stimulated great interest over the last 10 years, including more than 15 articles in Technometrics.

3. OPTIMAL DESIGN

The traditional motivation underlying the theory of optimal design is that experiments should be designed to achieve the most precise statistical inference possible. Kiefer (1981) stated that research work on optimal design arose in part as a reaction to earlier research on design, which emphasized attractive combinatoric properties rather than inferential properties. Design optimality was first considered by Smith (1918), and early work in the subject was done by Wald (1943), Hotelling (1944), and Elfving (1952). The major contributions to the area, however, were made by Kiefer (1958, 1959) and Kiefer and Wolfowitz (1959, 1960), who synthesized and greatly extended the previous work. Although the ideas of optimal design initially generated considerable controversy (see, for example, the discussion accompanying the paper by Kiefer 1959), they have since become well-established in the statistical literature. In some areas, such as the design of block

experiments, the use of optimal design theory is now accepted as a fundamental tool for comparing designs (see Section 9). In other areas, however, there is still disagreement over the applicability of optimal design theory (see, for example, the discussion in Section 6 on response surface designs).

Excellent reviews of research work on optimal design have appeared. For readers interested in the most recent developments in optimal design, we recommend the reviews by Atkinson (1982), Pazman (1980), and Ash and Hedayat (1978). The review by St. John and Draper (1975) provides a good introduction to the topic. The recent book by Silvey (1980) presents a concise summary of the classical results in optimal design theory, and the book by Fedorov (1972) is a valuable compendium of results.

The influence of optimal design has extended to almost all areas of experimental design, and it will be useful to review some of the most basic definitions and results because they will be needed in subsequent sections. To apply optimal design theory in practice requires a criterion for comparing experiments and an algorithm for optimizing the criterion over the set of possible experimental designs. We will define the most commonly used criteria here but will defer the consideration of algorithms to Section 4. The classical criteria are derived within the context of linear model theory in which it is assumed that the experimental data can be represented by the equation

$$Y_i = f(x_i)' \beta + \epsilon_i, \quad (3.1)$$

where Y_i is the measured response from the i th experimental run, x_i is a vector of predictor variables for the i th run, f is a vector of p functions that model how the response depends on x_i , β is a vector of p unknown parameters, and ϵ_i is the experimental error for the i th run.

A natural way to measure the quality of statistical inference with respect to a single parameter is in terms of the variance of the parameter estimate. If the errors are uncorrelated and have constant variance σ^2 , the variance-covariance matrix of the least squares estimator $\hat{\beta}$ is

$$\text{var}\{\hat{\beta}\} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}, \quad (3.2)$$

where \mathbf{X} is the $n \times p$ matrix whose i th row is $\mathbf{f}(\mathbf{x}_i)'$. We will limit our discussion here to the case where \mathbf{X} has full column rank. (Mathematically, the theory is not substantially different if \mathbf{X} is not of full rank.)

Another useful way to measure the quality of inference is in terms of the variance of the estimated response at \mathbf{x} , which, from (3.1), is given by

$$d(\mathbf{x}) = \sigma^2 \mathbf{f}(\mathbf{x})' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{f}(\mathbf{x}). \quad (3.3)$$

Both (3.2) and (3.3) depend on the experimental design only through the $p \times p$ matrix $(\mathbf{X}'\mathbf{X})^{-1}$, and suggest that a good experimental design will be one that makes this matrix small in some sense. Since there is no unique size ordering of the $p \times p$ matrices, various real-valued functionals have been suggested as measures of "smallness." The most popular of these optimality criteria are listed below:

1. D-Optimality - A design is said to be D-optimal if it minimizes $\det (\mathbf{X}'\mathbf{X})^{-1}$, where \det denotes determinant.
2. A-Optimality - A design is said to be A-optimal if it minimizes $\text{tr} (\mathbf{X}'\mathbf{X})^{-1}$, where tr denotes trace.
3. E-Optimality - A design is said to be E-optimal if it minimizes the maximal eigenvalue of $(\mathbf{X}'\mathbf{X})^{-1}$.
4. G-Optimality - A design is said to be G-optimal if it minimizes $\max d(\mathbf{x})$, where the maximum is taken over all possible vectors \mathbf{x} of predictor variables.

5. I_λ -Optimality - A design is said to be I_λ -optimal if it minimizes $\int d(\underline{x})\lambda(d\underline{x})$, where λ is a probability measure on the space of predictor variables. This criterion, which is sometimes called average integrated variance, also belongs to a more general class of L-optimality criteria discussed by Fedorov (1972).

One important result in optimal design theory is the general equivalence theorem (Kiefer and Wolfowitz 1960), which links D- and G-optimality. The theorem is phrased in terms of design measure, in which a design is represented by a probability measure on the predictor variable space. Thus, for example, a trial of n runs (an "exact" design) would be represented as a discrete measure with mass $1/n$ at each of the n points of the design. The concept of design measure is useful in studying optimal design theory from a mathematical point of view because it replaces a discrete optimization problem (finding the optimal "exact" design) with a continuous problem (finding the optimal design measure) which is often easier to solve. Although the solution to the continuous problem might, in theory, be a measure with infinitely many support points, Kiefer and Wolfowitz (1960) showed that solutions could always be limited to measures with finitely many support points; the value of the measure at each point would then give the optimal proportion of runs that should be made there. The General Equivalence Theorem states that among design measures ξ , the following three conditions are equivalent:

1. ξ^* is D-optimum.
2. ξ^* is G-optimum.
3. $\max d(\underline{x}; \xi^*) = p$.

The third condition, which provides a simple way to check whether a design is D- and G-optimal, is useful in constructing such designs.

4. COMPUTER-AIDED DESIGN OF EXPERIMENTS

Research on the use of computers in the design of experiments has been closely related to the increasing attention given to optimal design in the literature. As described in the previous section, the basic idea of optimal design is usually to choose a design that optimizes some inference criterion over the set of designs being considered. In practice, this optimization problem may be difficult or impossible to solve analytically. The first research done on using the computer as an essential aid in tackling this problem in experimental design was apparently that by Box and Hunter (1965a,b) on design for nonlinear models. The remainder of the present section will emphasize the use of computers in the design of linear regression and factorial experiments. We will discuss the two topics in turn.

4.1 Regression Experiments

Much research has been concerned with the development of constructive algorithms that can be used to find optimal or near-optimal designs. Initial work on the development of such algorithms focused on finding D- and G-optimal design measures (Wynn 1972, Fedorov 1972, Atwood 1973). These algorithms involve the construction of a sequence of design measures in which each succeeding measure is a convex combination of the current measure and a point mass whose location is chosen with the aid of the third condition of the General Equivalence Theorem stated in Section 3. General conditions for such algorithms to converge to an optimal design measure were given by Wu and Wynn (1978).

Computer-aided design of regression experiments was stimulated by the desire to achieve exact n -run optimal designs. In some cases good designs can be found directly from an optimal design measure by spreading out the runs to

approximate the optimal allocation. For designs with a small number of runs or models with many parameters, however, this strategy may be difficult to implement or may lead to designs that are quite inefficient.

Improvements in computer technology have made it possible to adopt an alternative scheme: developing computer programs to directly find exact n-run optimal designs. The most popular computer algorithm developed to date, known as DETMAX, was originated by Mitchell (1974a,b) to find D-optimum designs. This program requires the user to specify the model and the number of experimental runs (n), and to list all the possible design points for the experiment. An initial n-run starting design may be supplied by the user or generated by the program. The program then seeks to maximize $\det(\mathbf{X}'\mathbf{X})$, which is equivalent to minimizing $\det(\mathbf{X}'\mathbf{X})^{-1}$, by adding and deleting design points until a convergence criterion is satisfied. The choice of which point to add or delete at each step is made so that $\det(\mathbf{X}'\mathbf{X})$ is maximized among all possibilities. Galil and Kiefer (1980b) showed that this criterion will always add a design point at which the variance of the estimated response (equation 3.3) is greatest; this property is related to the result of the general equivalence theorem that the variance of the estimated response for an optimal design measure obtains its maximum value at each of the design points. The DETMAX program allows the possibility of "excursions," in which several points are added and then several points deleted, in the hope of avoiding local maxima. Mitchell (1974a) also recommended that a number of different starting designs be used because no one starting design is guaranteed to lead to an optimum design. Several characteristics are calculated for designs at or near the D-optimal design and these properties may be used as additional bases of comparison.

Mitchell (1974b) used the DETMAX program to tabulate designs for first-order regression models with up to nine factors and a variety of sample sizes. Mitchell and Bayne (1978) used the program to find fractions of three-level factorial designs for models including some two-factor interactions and for models including two-factor interactions and pure quadratic terms.

Galil and Kiefer (1980b) developed useful modifications to DETMAX, which led to a substantial reduction in the amount of time needed to search for an optimal design and in the amount of computer space required by the program. They also proposed a systematic method for generating an initial design. The reduction in time is quite important because it allows many more starting designs to be used for a fixed computer budget, thereby increasing the chance of finding an optimal exact design. The space-saving methods make it possible to study larger problems. Galil and Kiefer also studied in detail the problem of quadratic regression for designs that are fractions of three-level factorials and tabled the best designs found by the modified DETMAX algorithm.

A new program developed by Welch (1982) takes advantage of the branch-and-bound optimization strategy. This program is more powerful than DETMAX in that it is assured to find all possible n -run optimal designs for a given model and a specified set of possible design points. It is not clear, however, what additional cost in Central Processing Unit (CPU) time may be involved. Welch also considered in detail the problem of quadratic regression with three-level factors and proved that some of the designs tabled by Galil and Kiefer (1980b) were, in fact, D-optimal.

Snee and Marquardt (1974) considered the special problem of optimal design for mixture experiments (see Section 7). Their XVERT program was designed to find extreme vertices of the design region and to calculate

several optimality criteria for a variety of extreme vertex designs. Snee (1979) described the CONSIM algorithm for finding extreme vertices and centroids of mixture design regions and recommended that it be combined with XVERT (when there are at most four mixture components) or with DETMAX (when there are five or more components) to generate experimental designs. Nigam, Gupta, and Gupta (1983) proposed a modified version of the XVERT algorithm for finding extreme vertices of mixture design regions; the modified algorithm involves less computational effort than does XVERT but the authors found that there was little loss of efficiency.

One of the common features of the above papers is their use of a design region with a finite number of possible design points, rather than a continuous region with infinitely many points. The primary reason for limiting the algorithms to finite design spaces is to simplify the task of selecting what point to add to an existing design. The benefit of each candidate point can be computed and the best point is then selected. The use of a finite design region is reasonable on practical grounds even when some of the factors are continuous, quantitative variables because an experimenter's ability to exactly fix the levels of quantitative factors in an experiment is limited. If, however, there are many possible settings for each quantitative factor, so that the number of design points, although finite, is quite large, a more efficient strategy may be to treat the design region as though it were continuous. To choose the new design point from a continuous region, a functional optimization algorithm must be implemented and the success of the design algorithm will depend, at least in part, on the ability of the optimization algorithm to find the best point to add at each iteration.

Cook and Nachtsheim (1980) used Powell's (1964) conjugate direction method to maximize $\det(\mathbf{X}'\mathbf{X})$ at each iteration and compared several computer design algorithms. Not surprisingly, they found that the best results were obtained by those algorithms that required the most CPU time. They concluded that DETMAX (without Galil and Kiefer's modifications) gave good results relative to the amount of CPU time it required.

Evans (1979) presented a simple computer algorithm for augmenting an existing experimental design by a fixed number of runs so that the combined design would be D-optimal. His algorithm called for simultaneously choosing all the new design points, rather than the sequential selection characteristic of other methods, and used a modified version of Nelder and Mead's (1965) simplex method to maximize $\det(\mathbf{X}'\mathbf{X})$. Both DETMAX and Welch's algorithm also allow the user to augment an existing design, although within the framework of a finite design space.

Johnson and Nachtsheim (1983) studied several problems in the construction of designs on continuous, convex design spaces. They concluded that Evans's approach of simultaneously searching for all the new points to be added to an existing design offered little improvement over sequential search procedures. They compared several optimization algorithms for choosing the point that maximizes $\det(\mathbf{X}'\mathbf{X})$ and found that Powell's (1964) algorithm gave the best results. Finally, they found that Galil and Kiefer's (1980b) method of generating an initial design was quite successful.

One of the first articles on computer-aided design of regression experiments took an approach quite different from those described previously. Kennard and Stone (1969) argued that a good design should cover the design space as uniformly as possible. They developed the CADEX algorithm to achieve this goal by sequentially choosing that point furthest from the current design

points. They favored the use of this uniform coverage criterion because it does not require the assumption of any particular model such as (3.1) for the response and because, when several response variables are measured, the same design will be appropriate for each one.

4.2 Factorial Experiments

Computer algorithms have also been developed to aid in the design of factorial experiments. Patterson (1976) described the DSIGN program, which produces designs for factors at any number of levels with a variety of blocking structures, including Latin squares and split plots, according to a generating design key supplied by the user. The design key specifies the plot aliases of the main effects of treatment factors. Bailey, Gilchrist, and Patterson (1977) and Patterson and Bailey (1978) described the use of design keys in identifying confounding patterns and in constructing designs. The designs produced by the DSIGN program are compared on the basis of their confounding patterns, rather than any of the formal optimality criteria mentioned earlier.

Jones and Eccleston (1980) described a computer algorithm for the generation of optimal block designs. As an optimality criterion, they proposed minimizing the weighted sum of the variances of a set of treatment contrasts, which is similar to the criterion for A-optimality. This criterion depends on two characteristics of the design: the replication numbers, which state how many times each treatment will be used, and the set of concurrences, which gives the number of times each pair of treatments occurs in the same block. The algorithm determines the replication numbers and concurrences in two separate stages, known as exchange and interchange. Beginning with an initial design that specifies the treatments assigned to each block, the

exchange procedure locates runs that contribute little to the optimality criterion and seeks to find different treatments for those runs that will make a greater contribution. Thus the exchange procedure alters the initial replication numbers and the set of concurrences. The interchange procedure then seeks to improve the optimality criterion by switching the block assignments of pairs of treatments (e.g., the blocks ABC, DEF might be changed to ABF, DEC by interchanging treatments C and F). Thus the interchange procedure does not affect the replication numbers but does change the set of concurrences. Eccleston and Jones (1980) extended the exchange-interchange algorithm to designs for the elimination of both row and column effects.

Wu (1981a) presented a computer algorithm for assigning experimental units to different treatments when categorical covariate information is available for each unit. The algorithm is designed to balance the covariates across the different treatments, yet is surprisingly simple and requires no matrix inversion.

5. DESIGN ROBUSTNESS

Box (1953) introduced the word "robust" in the statistical literature to describe procedures that give good results even though there might be violations in the assumptions upon which these procedures are based. Following up a line of research initiated by Pearson (1931), Box (1953) examined the effect on the analysis of variance and on Bartlett's test of departures from normality, an assumption underlying both procedures. Pearson (1931) had discovered that the analysis of variance is robust to such violations of assumption, but suggested that his conclusion would not be valid for comparing estimates of variance based on independent samples. Box (1953)

found that, indeed, Bartlett's test is quite sensitive to departures from normality. This result led him to observe that the use of Bartlett's test as a preliminary to the analysis of variance -- a practice recommended by some statisticians at the time -- was "rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" (p. 333).

The examination of standard statistical techniques to determine their sensitivity to assumptions and the development of new techniques that are less sensitive have been focal points of statistical research in the last two decades (see Huber 1981). Experimental design is an area in which it is particularly compelling to investigate questions of robustness because a researcher's assumptions about the experimental process are often crucial in determining the design. Moreover, the design must be chosen before the data are collected and so cannot be discarded if the data indicate that the assumptions are seriously incorrect. (By contrast, techniques for data analysis may be replaced by other alternatives if their use is contraindicated by the observed data.) Thus it is important to examine experimental designs for sensitivity to assumptions. Interest in design robustness, therefore, should come as no surprise; if anything, we are surprised that this topic has not attracted greater attention.

The assumption that underlies most research work in experimental design is that the experiment can be adequately described by an equation of the form:

$$\text{response} = \text{model} + \text{error}, \quad (5.1)$$

where the model states the effect of the predictor variables on the response variable and the error describes the general form of departures from the model. Experimenters frequently have tentative models in mind, either on the basis of theoretical considerations or on the belief that a simple empirical

model will be adequate, at least over the current range of experimentation. It is unlikely, however, that the experimenter will be absolutely certain that any tentatively entertained model will be adequate, and design strategies that fail to take this uncertainty into account must be viewed with some skepticism. In particular, designs derived using the optimality criteria discussed in Section 3 are known to depend quite critically on the particular model that is assumed. These designs tend to concentrate all the experimental runs on a small number of design points and are ideally suited to estimating the coefficients of the assumed model, but they provide little or no ability to check for lack of fit. Assumptions about the error component in (5.1) are typically characterized in terms of a probability distribution and are also subject to uncertainty. The research work reviewed in this section concerns the consequences for experimental design of misspecifying the form of the model or the error.

Two different, but complementary, approaches have been proposed for planning experiments in the face of model uncertainty. The first approach has sought designs that will yield reasonable results for the proposed model even though it is known to be inexact. We call these designs "model-robust designs" and discuss them in Section 5.1. The second approach has focused on developing designs that facilitate improvement of the proposed model by trying to highlight suspected inadequacies. We call these designs "model-sensitive designs" and discuss them in Section 5.2.

Another line of research in design robustness concerns the implications for experimental design of inaccurate assumptions about the error rather than the model. We call these "error-robust designs" and discuss them in Section 5.3.

5.1 Model-Robust Designs

Box and Draper (1959) were the first authors to consider in depth the effect of model misspecification on experimental design. They criticized the common optimality criteria defined in Section 3 for implicitly assuming that the proposed model is exactly correct. They argued that a more appropriate criterion for comparing experimental designs is the average mean squared error (J) over a region of interest R, which is contained in the total experimental region:

$$J = (n/\sigma^2\Omega) \int_R E\{[\hat{g}(\underline{x}) - g(\underline{x})]^2\} d\underline{x}, \quad (5.2)$$

where $g(\underline{x})$ is the true response function, $\hat{g}(\underline{x})$ is the least squares estimate of $g(\underline{x})$, and $\Omega = \int_R d\underline{x}$. This expression can be decomposed as the sum of a bias component and a variance component:

$$J = (n/\sigma^2\Omega) \left(\int_R [E\{\hat{g}(\underline{x})\} - g(\underline{x})]^2 d\underline{x} + \int_R \text{Var}\{\hat{g}(\underline{x})\} d\underline{x} \right), \quad (5.3)$$

respectively. Box and Draper (1959) considered, in particular, the effect of assuming a first-degree polynomial regression model when the true model is a second-degree polynomial. They found that the designs that minimized average mean squared error were similar to those that minimized the bias component alone, but were quite different from those that minimized the variance component. Thus their "minimum bias" designs differed markedly from those implied by the traditional optimality criteria, which consider only functions of the variance. The "minimum bias" designs could be found by choosing the design points in such a way that specified moments of the design matched those of a uniform probability distribution on the region of interest.

Box and Draper (1963) and Huber (1975) reached similar conclusions. Box and Draper (1963) extended the work discussed above by studying the situation in which the assumed model is quadratic but the true model is cubic. Huber

(1975) investigated the sensitivity of optimal design to model misspecification by conducting a minimax analysis. For a given design, he determined what true response function would lead to the greatest mean squared error. Huber found that optimal designs based on first-degree polynomial regression models could be subject to considerable bias from quadratic terms.

Kussmaul (1969) investigated the effect of model misspecification in simple polynomial regression. In particular, he was concerned with the fact that classical optimal designs tend to concentrate all the experimental runs at a small number of design points. For example, the D- and G-optimal design measure for estimating a polynomial model of degree j locates experimental runs at exactly $j + 1$ distinct levels of the predictor variable. This design can provide no indication that a higher degree polynomial may be needed. Kussmaul suggested that this problem might be overcome by using the G-optimal design for a polynomial model of degree k , with k slightly greater than j . He concluded that the loss of efficiency using this design strategy, with respect to the G-optimality criterion, was quite small and was more than offset by the added protection of being able to fit a polynomial of higher degree, if necessary.

Lauter (1974) considered the general problem of optimal design when the form of the true response function is unknown but assumed to belong to a specified class of linear models. She proposed extending the common optimality criteria for an exactly assumed model to the broader class of models by using different forms of averaging with respect to a weighting measure on the class of models. The resulting designs are not, in general, optimal for any of the models, but she claims they should be reasonably efficient for all models considered likely. Cook and Nachtsheim (1982) applied Lauter's general approach to polynomial regression. They assumed that

a low-degree polynomial would probably be adequate to approximate the true response function, but that a polynomial of higher degree might be necessary. Designs were compared on the basis of average inefficiency, where the inefficiency for a polynomial of fixed degree was calculated by comparison with the best design for that degree. The average was weighted to reflect the assumption that a low-degree polynomial would probably be adequate by placing most of the weight on those models.

Several authors have considered modifying the basic model to include possible effects of model inadequacy. O'Hagan (1978) postulated a Bayesian model in which β in (3.1) is replaced by $\beta(\mathbf{x})$. The dependence of β on \mathbf{x} is characterized by a prior probability distribution that reflects beliefs about the likely smoothness and stability of the true response function. For this model, he found that a design criterion based on posterior variance favored placing more points near the center of the design region when compared with constructing designs on the basis of criteria like D-optimality. The discussion to O'Hagan's paper provides a lively introduction to different approaches to model robustness in experimental design. Smith and Verdinelli (1980) adopted a hierarchical Bayesian model of the form analyzed by Lindley and Smith (1972). The hierarchical structure incorporates a particular model such as a low-degree polynomial but also reflects the degree to which the experimenter is confident that the polynomial model is adequate. They examined the allocation of runs to a fixed set of doses in a dose-response experiment and found that perfect confidence in an assumed polynomial model led to the D-optimum allocation for that model. As confidence in the model decreased, however, the allocation changed smoothly to an even distribution of runs among the doses. Pesotchinsky (1982) studied the implications for design of the "approximately linear" regression model of Sacks and Ylvisaker (1978),

in which the response function is assumed to differ from a first-degree polynomial by, at most, a fixed convex function. He found that the designs depended on the form of the fixed function, its magnitude relative to experimental error, and the sample size. Although Pesotchinsky's approach to incorporating model inadequacy is quite different from those of O'Hagan and of Smith and Verdinelli, who used Bayesian models, the magnitude of potential bias relative to experimental error proved to be a key parameter in all three papers.

Wu (1981b) considered a different type of robustness -- the possibility that a simple additive model for a block design would be violated by the addition of fixed, but unknown, unit effects. Using a minimax criterion to study the sensitivity of different design strategies, he concluded that randomized assignment of the units to the treatments was the best way to obtain designs robust to the contaminating unit effects. Wu's results have been generalized by Li (1983).

5.2 Model-Sensitive Designs

The research reviewed above was motivated by the desire to make the analysis of the experiment insensitive, or robust, to possible uncertainties or inaccuracies in the specification of the model. The experimenter's primary interest, however, may be to highlight the uncertainties and inaccuracies in order to modify or refine the model initially entertained; the experimenter will then require a design that is sensitive to the differences between alternative models. Model-robust and model-sensitive designs are quite similar to one another and share much common ground because the essential idea behind both concepts is that tentatively proposed models are never exact.

Some work on model-sensitive designs has been motivated by studies in which nonlinear models are used (see Sect. 10), although the theory developed is equally applicable to linear models. The experimenter may be able to list a set of plausible nonlinear models, perhaps through knowledge of the underlying experimental mechanism or previous experience with similar experiments. Experiments are then designed whose primary purpose is to discriminate among candidate models. Hunter and Reiner (1965), Box and Hill (1967), Atkinson and Fedorov (1975a,b), and Atkinson (1981) discussed nonlinear models. Atkinson and Cox (1974) discussed linear models. These techniques have usually been referred to as model discrimination designs; they are a special case of what we define here as model-sensitive designs.

Hill, Hunter, and Wichern (1968) suggested the use of a design criterion that simultaneously takes into account the needs of model discrimination and parameter estimation; they illustrated its use with nonlinear models. Atkinson (1972) proposed a criterion for the design of linear regression experiments that have the joint aim of estimating parameters in a tentatively assumed model and of testing for inadequacy of that model. He considered models of the form:

$$y_i = f_1(x_i)' \beta_1 + f_2(x_i)' \beta_2 + \varepsilon_i, \quad (5.4)$$

where $f_1(x_i)$ corresponds to the tentatively assumed model and $f_2(x_i)$ corresponds to those additional terms thought most likely to induce bias. For example, f_1 might be a first-degree polynomial and f_2 might be a function containing quadratic terms. Selecting from standard classes such as factorials and central composite designs, Atkinson found experimental plans that compromise between the two goals.

Jones and Mitchell (1978) studied designs whose primary purpose is to detect inadequacy of a tentatively assumed linear model in the direction of a specific alternative model. They also considered models of the form (5.4),

but used criteria different from Atkinson's. One major difference is that Jones and Mitchell's designs depend on the unknown parameter vector β_2 in (5.4), while Atkinson's do not. Jones and Mitchell suggested two methods to overcome this dependence, both of which are related to a design criterion proposed by Atkinson and Fedorov (1975a,b).

A related approach is that of Stigler (1971), who proposed the idea of restricted D- and G-optimal designs for polynomial regression, in which the optimal design for an assumed polynomial model would be found subject to the restriction that the design should make it possible to estimate a higher-degree model with some minimal level of precision. This approach leads to a range of designs that compromise between the unrestricted optimal designs, at one extreme, and designs that provide maximum power for testing that the higher degree coefficients are all zero, at the other extreme. Studden (1982) described an elegant method for constructing such designs.

Morris and Mitchell (1983) studied the special case of designing two-level multifactor designs that are sensitive to detecting interactions among the factors. They recommended a sequential approach in which screening for interactions is done at the earliest possible stage in an experiment and then forms a basis for planning subsequent runs. They proposed a design criterion and gave rules for the construction of designs optimizing the criterion. Their methods make it possible to screen for interactions with only a small number of experimental runs.

Sometimes the degree of model uncertainty is so great that it becomes impractical to pursue the methods described above. The task of specifying all possible models or classes of models and then optimizing a design criterion may be unmanageable because there are too many models to deal with. In some applications in chemical kinetics, for example, more than 100 candidate models

can be listed, all of them nonlinear. Sometimes an experimenter is confronted with the opposite situation: it is difficult to specify even a single model. In either of these circumstances it may be best to proceed by putting forward one model on a tentative basis, even though it is almost certain to be incorrect in some important respects. The experimenter might then wish to employ an efficient experimental plan whose purpose is to provide data that will, with the greatest sensitivity possible, reveal the inadequacies of the initial model so that it can be modified or replaced altogether. A desirable property of such a design is to display shortcomings in the model in a manner that will best help the investigator to create a better model. The procedure can be repeated with subsequent models. Unfortunately, the literature on useful methods of model-building along this line is limited; see, for example, Box and Hunter (1962), Hunter and Mezaki (1964), Draper and Herzberg (1971), and Box and Draper (1982).

5.3 Error-Robust Designs

The error terms in (5.1) are typically assumed to be independent and identically distributed. Further, the common distribution is often assumed to be a normal distribution. This section will review research in design robustness that has considered violations in the assumptions concerning the distribution of the error terms.

Box and Draper (1975) and Huber (1975) studied the possible effects of outliers on experimental design for linear models. Huber suggested that resistance to outliers in the error distribution could be achieved by avoiding outlying points in the experimental design. To accomplish the latter goal, he recommended using designs for which the diagonal elements of the "hat" matrix $H = X(X'X)^{-1}X'$ are well below unity (see Hoaglin and Welsch 1978 for a

discussion of the "hat" matrix and its relationship to outliers). Box and Draper (1975) showed that the effect of one or more wild observations on the vector of predicted values is proportional to $\sum h_{ii}^2$, where h_{ii} is the i th diagonal element of the matrix H defined above. This sum is minimized if $h_{ii} = p/n$ for all i , so that Huber's recommendation may be interpreted as adding to the Box-Draper criterion a requirement that p/n not be too large. Draper and Herzberg (1979) extended the work of Box and Draper (1975) by studying the effect of outliers on mean squared error when the model is a polynomial of low degree but is subject to bias from terms of higher degree.

Herzberg and Andrews (1976) and Andrews and Herzberg (1979) considered the possibility that some observations would be missing altogether or would be so extreme that they would be entirely discarded from the analysis. They proposed several measures of robustness against such occurrences, such as the probability that the "effective" X matrix (i.e., the X matrix for the remaining points) would not have full rank and the expected value of the D-criterion for the design, where the relevant probability distribution for these calculations is that which specifies the probabilities that the planned observations will actually be usable in subsequent analysis. They found that some conventional optimal designs are not robust under these criteria, which tend to favor designs with some repeated points.

Another standard assumption is that the error terms ϵ_i are stochastically independent. Sacks and Ylvisaker (1966) studied the problem of designing regression experiments when the errors are correlated, as might happen if the observations are realizations of a time series. They derived asymptotic characterizations of optimal designs, which call for taking all the observations at distinct points. These designs differ from the optimal designs derived under the assumption of independence, which tend to replicate

many observations at a small number of design points. These results were generalized by the authors in a later paper (Sacks and Ylvisaker 1968) and by Wahba (1971). Eubank, Smith, and Smith (1981, 1982) have proved some uniqueness results for these designs.

A different approach to the problem of correlated errors has been explored by Rickel and Herzberg (1979) and Rickel, Herzberg, and Schilling (1981). They developed asymptotic theory and numerical results for a model in which the extent of the correlation among the errors is assumed to decrease with the sample size (as might occur, for example, if additional observations were spread over a wider interval). For situations in which the errors are assumed to follow a first-order autoregressive process, the authors derived designs for estimating location and for fitting simple linear regression models. These designs are described exactly, whereas the papers mentioned in the previous paragraph gave only complicated characterizations of designs.

Another approach to time dependence is to consider the possibility that the sequential order of the experimental runs will affect the results through a polynomial trend. Such a trend can then be included in the model component of (5.1), rather than in the error component, and it is possible to develop designs that are orthogonal to the trend. This approach was first explored by Daniel and Wilcoxon (1966) in the context of factorial designs and was extended by Joiner and Campbell (1976). Bradley and Yeh (1980) developed a theory for trend-free block designs.

6. RESPONSE SURFACE DESIGNS

Response surface methodology was developed by Box and his colleagues at Imperial Chemical Industries to explore relationships such as those between the yield of a chemical process and the pertinent process variables (Box and Wilson 1951, Box 1954, Box and Youle 1955). In its usual form, response

surface methodology exploits simple empirical models such as low-degree polynomials to approximate the relationship between a response variable and a set of input variables over a current region of interest.

A key intellectual insight in the development of response surface methodology was the realization that in chemistry, engineering, and physics, experimental data are often available for analysis much more rapidly than in agriculture. Thus an efficient way to organize experimental programs in chemistry, engineering, and physics is to adopt a sequential strategy in which the experiment proceeds in stages, with each stage designed in the light of results obtained from earlier runs. The classic factorial designs formed a basis for the construction of the design at each stage, but typically the designs were smaller than those used in agriculture. A second important insight was that the experimental variables in the chemical, physical, and engineering sciences are frequently quantitative (continuous), whereas the variables in agricultural experiments are often qualitative (categorical); this led to the useful idea of rotatable designs, proposed by Box and Hunter (1957), which seeks designs for which the variance of estimated responses is constant on spherical shells in the region of interest. Once these differences had been recognized, the way was open to develop new, more efficient experimental design strategies that took advantage of them.

Since its introduction in the early 1950's, response surface methodology has become an accepted and widely used set of concepts and techniques. Chapter 11 of Davies (1954), Chapter 8A of Cochran and Cox (1957), Chapter 10 of John (1971), and the book by Myers (1976) contain explanations of the basic ideas of response surface methodology, including both the design of response surface experiments and the estimation and interpretation of the fitted surface. An introduction to the subject at a more elementary level is given in Chapter 15 of Box, Hunter, and Hunter (1978). The review article by Hill

and Hunter (1966) contains over one hundred references. Finally, a definitive book by Box and Draper (1984) is scheduled for publication.

One of the important applications of response surface methods, in a simplified form, has been to improve the performance of existing industrial processes by systematically varying process variables and gathering data while the processes operate, without upsetting normal production. This use of response surface methods for process improvement is known as evolutionary operation (EVOP). EVOP is an aggressive management strategy in which better ways of operating a process are actively sought rather than accidentally discovered. Box and Draper (1969) explain the fundamental principles and methods of EVOP and discuss how an EVOP program can be implemented. Spendley, Hext, and Himsworth (1962) proposed an alternative design scheme, known as simplex EVOP, in which k process variables are placed on a simplex with $k + 1$ vertices. Hahn and Dershowitz (1974) discussed some of the practical issues that should be considered in using EVOP and reported the results of a survey indicating that EVOP is not being used in industry as widely as it could be.

Much of the statistical work on response surface design in recent years has concerned the use of optimal design theory for response surface studies. Some authors have advocated the application of the precepts of optimal design theory to derive response surface designs. Others, however, have questioned the applicability of optimal design theory to response surface experiments. Typical of the former group is the series of papers by Galil and Kiefer (1977a,b, 1979), in which they derived optimal designs for quadratic and cubic polynomial response surface models when the domain of the predictor variables is assumed to be a k -dimensional cube or sphere. Designs were derived for a family of optimality criteria that includes A-, D-, and E-optimality. The

efficiency of the designs was compared for these and other criteria in the family. Pesotchinsky (1978) gave expanded results for quadratic models. Lucas (1976) compared a variety of designs for quadratic models on the basis of the D- and G-optimality criteria. Some of the computer-derived optimal designs reported in Section 4 (Mitchell 1974b, Mitchell and Bayne 1978, Galil and Kiefer 1980b, Welch 1982) also apply to response surface experiments.

Criticism of the use of optimal design theory for response surface experiments has focused on several issues. One of the major arguments against the use of optimal design theory is the need to specify a model for the response function, coupled with the fact that optimal designs are frequently quite sensitive to the form of the model. This concern stimulated much of the research recounted in Section 5 to achieve experimental designs that are robust to the choice of the model. Box and Draper (1959) and Box (1982) argued that to assume that a linear model such as (3.1) exactly represents the true response function is especially troubling in response surface studies, where the linear model is never intended to be more than a reasonable local approximation. Hence, they advised that the possible effects of bias be considered in choosing a design.

Box (1982) voiced further criticism of the use of optimal design theory in response surface studies. Optimal designs for a particular model and criterion are found by optimizing the criterion over the set of possible designs, which is typically defined in terms of a prescribed region of experimentation, within which the predictor variables must be set. A common assumption is that the region of experimentation is a simple geometric body, such as a hypercube or hypersphere, or can be transformed to such a region by centering and scaling the predictor variables. One of the characteristics of the optimal designs found in the papers mentioned above is that many

experimental runs are placed at the extreme limits of the region. Box (1982) observed that, in most response surface experiments, the region of possible experimentation is not precisely known; moreover, as design points are moved further away from one another, the effect of bias on any simple approximating model is likely to become increasingly severe. Thus the tendency of the optimal designs to concentrate many runs at the extremes of the design region must be viewed with some trepidation, especially in the context of response surface experiments. Similar criticisms were also stated by O'Hagan (1978) and helped stimulate his Bayesian approach to design.

Other topics related to the design of response surface experiments have recently been studied. Draper (1982) discussed several methods for choosing the number of center points in designs for quadratic response surface models. Hader and Park (1978) proposed the concept of slope-rotatable designs, for which the variance of the first partial derivatives of the estimated response function would be constant on spherical shells centered at the origin. Slope-rotatability might be a desirable property if the primary purpose of the experiment is to estimate the slope of the response surface and there is equal interest in estimating the slope in all directions from the center of the experimental region. This work is related to that by Box and Hunter (1957), who proposed the use of rotatable designs, which have the property that the variance of the estimated response is constant on spherical shells. Box and Draper (1982) discussed several measures of lack of fit for response surface designs, and how the fit might be improved by power transformations of the predictor variables. Box and Draper (1980) discussed a geometric interpretation for the variance of the difference between two estimated responses and gave results for quadratic and cubic rotatable designs.

7. MIXTURE DESIGNS

In some experimental situations the response depends on the relative amounts of the predictor variables, but not on the absolute amounts. Typical examples would be car mileage as a function of the proportions of components blended into gasoline or the strength of an alloy as a function of the fractional amounts of constituent metals. The special nature of these experiments, known as mixture experiments, can be expressed in the following set of constraints: if X_1, \dots, X_k denote the k predictor variables, measured as proportions, then for each experimental run we must have:

$$0 \leq X_j \leq 1 \quad \text{for all } j, \text{ and } \sum_{j=1}^k X_j = 1. \quad (7.1)$$

This constraint presents some special problems for experimental design and statistical modeling because any model which contains linear terms in all the predictor variables and a constant term will be overparameterized; the sum of the k linear coefficients must be confounded with the constant term due to the constraint (7.1). Although the particular concern of mixture experiments is summarized by the constraint (7.1), the theory can be applied more generally to any problem in which there exist one or more linear constraints on the predictor variables.

A fortuitous circumstance in the development of statistical procedures for mixture experiments was Scheffé's work as a consultant for Chevron Research Corporation. Investigators there who were working on problems related to gasoline blending asked him for advice on the design of experiments in which the relative proportions of particular formulations were to be varied. These problems stimulated Scheffé to undertake the first systematic statistical study of mixture experiments. Scheffé (1958) introduced a family

of models for mixture problems and proposed the class of lattice designs, which place experimental runs on a uniform lattice of points, enabling the experimenter to explore response variables throughout the entire design simplex. Scheffé (1963) proposed simplex-centroid designs, in which runs are made using mixtures that have equal proportions of some subset of the components. One stimulus for the simplex-centroid designs was to correct a weakness of the lattice designs -- their tendency to use many experimental mixtures that involve only two components, even when the number of components is large.

A difficulty encountered in many mixture experiments is that some of the components are subject to upper or lower bounds. Such bounds can produce design regions with odd shapes for which it is impossible to use the designs mentioned above. McLean and Anderson (1966) proposed solving this problem by making experimental runs at the extreme points and various centroids of the constrained design region. These plans are known as extreme vertices designs and, as with Scheffé's designs, they allow exploration of the entire experimental region.

Much of the subsequent work on designs for mixture experiments has roots in the models and designs of Scheffé and in the extreme vertices designs of McLean and Anderson. The most complete reference for mixture problems is Cornell (1981), which is a readable introduction to the topic and also discusses much of the most recent research. Reviews of this work are also available (Cornell 1973, 1979). Although first proposed for gasoline blending experiments, mixture designs have been applied in a wide variety of situations. We believe that, in the future, applications of mixture designs will extend to many new fields as more experimenters become aware of their usefulness. The most interesting example that has come to our attention

recently is the use of mixture designs at a cooperative in France for blending different wines to produce a table wine. Previously, only blends that qualified as ordinary table wines had been produced. However, the designs succeeded in identifying a blend that received the higher grade of vin délimité de qualité supérieure, allowing the cooperative to sell it at a premium price. In the remainder of this section, we will review the most recently published research on mixture designs.

The development of computer programs to assist in selecting experimental runs has been a particular concern in mixture problems, especially when there are additional bounds on some of the components. The extreme vertices designs have been quite popular here, so that most computer programs proceed in two stages: first, the extreme vertices and centroids of the constrained design region are identified; then, a design optimality criterion is used to select the vertices and centroids that will be included in the design. For specific references, we refer the reader to our discussion of these programs in Section 4 on computer-aided design.

A common situation in mixture experiments is that, in addition to k mixture variables, there are some process variables that are not subject to the constraint (7.1). Experimental designs for these problems must specify settings for both the mixture variables and the process variables. Hare (1979) generated designs by restricting the mixture variables to a cuboidal subset contained within the simplex defined by (7.1) and then crossing that subset with a cuboidal region for the process variables. A different approach to designing experiments with both mixture and process variables was taken by Vuchkov, Damgaliev, and Yontchev (1981). They used a sequential procedure to generate quadratic designs with high efficiency in terms of the D-optimality criterion.

When the experimenter wishes to explore only a limited region of the design simplex, an alternative approach that we feel deserves further attention is to replace the k linearly dependent mixture components by $k - 1$ linear functions of the components, often called pseudo-components. By treating the pseudo-components as the independent variables in the experiment, any standard response surface design that fits inside the simplex can be used. No special consideration is necessary for process variables: they can be included as additional variables in the response surface design. Box and Gardner (1966) proposed a similar idea. Their projection designs were defined by taking standard designs like two-level factorials and adjusting them to meet one or more linear constraints like (7.1). A disadvantage of the projection and pseudo-component designs is that they are typically not symmetric with respect to the original mixture components.

Cornell and Khuri (1979) proposed designs for three-component mixture models with the property that the variance of the predicted response is constant on concentric triangles about the centroid of the simplex. This idea is analogous to the concept of rotatable designs in response surface studies (Box and Hunter 1957). The designs are constructed by performing a nonlinear transformation of the coordinate system and then applying the theory of rotatable designs. Piepel (1983) offered guidelines to check the consistency of linear constraints used to restrict the region of experimentation.

8. FACTORIAL DESIGNS

Factorial designs, first developed by Fisher and Yates at Rothamsted, are one of the major contributions of statistical insight into experimental design. Their essential feature, the simultaneous study of several factors, is a marked departure from the common idea that experimenters should vary only

one factor at a time. As Fisher (1926) observed, factorial designs offer many advantages: each experimental run gives information on several factors, not just one; the experiment yields as much information about each factor as though it alone had been varied; valuable additional information is available through the ability to check for possible interactions among the factors; and in the event that no interactions are found, there is a much broader base for generalizing conclusions on the main effect of a factor, since the effect has been observed in a variety of experimental conditions.

A further advance was the introduction by Finney (1945) of fractional factorial designs. These designs allow experimenters to study the main effects and low-order interactions of several factors in far fewer runs than required to complete the full factorial designs by sacrificing the ability to estimate high-order interactions. Fractional factorial designs thus offer great economy of time and resources when, as is often the case, high-order interactions are negligible. Plackett and Burman (1946) described a useful class of highly fractionated orthogonal designs, in which the main effects of $n - 1$ two-level factors are estimated using just n runs. Box and Hunter (1961a,b) described in detail the theory and application of 2^{k-P} fractional factorial designs. For experiments in which some factors are used at more levels than others, Addelman and Kempthorne (1961) and Addelman (1962) presented a simple technique for deriving designs that give orthogonal estimates of main effects. The important contributions that factorial and fractional factorial designs can make to experimentation in the chemical, physical, and engineering sciences were clearly evident to the initial editors of Technometrics: many articles described these designs and illustrated their usefulness.

The most commonly used factorial designs can be found in most books on experimental design. John (1971) is an excellent source for the factorial designs used most often in practice: two- and three-level factorials and fractional factorials, including Plackett-Burman designs, main effect plans, and some asymmetric factorials (i.e., designs in which not all factors have the same number of levels). John's book is directed toward readers with some mathematical and statistical sophistication. Daniel (1976) also presents many useful factorial plans and describes a number of interesting applications; the level is less theoretical than John (1971), but some background in statistics is necessary. At a more elementary level, the book by Box, Hunter, and Hunter (1978) describes two-level factorial and fractional factorial designs in considerable detail. A good source for asymmetric factorial plans is the book by Cochran and Cox (1957), which also covers the topics mentioned immediately above, although with a bias toward agricultural examples and terminology. Davies (1954) also lists many useful factorial designs. Raktoe, Hedayat, and Federer (1981) is a concise but comprehensive treatise on the mathematical theory underlying factorial designs, with only a limited emphasis on applications.

Recent research on factorial designs has considered several problems, including incomplete factorials, weighing designs, screening designs, asymmetric factorials, and blocking schemes. A brief review follows.

John (1979) and Smith and Schmoyer (1982) both considered the effect on two-level factorial designs of incomplete replication. John showed that losing a single observation from a 2^k factorial experiment could double the variance of some of the estimated factor effects. He also examined the effect of missing observations on design resolution for 2^{k-p} experiments. Smith and Schmoyer investigated the consequences for a two-level factorial of

terminating the experiment prior to completing all 2^{k-p} runs in the original plan. Such a situation might arise due to equipment failure or to a conscious decision to cease experimentation, and it is especially relevant to the physical sciences, where experiments are often run sequentially (as opposed to the simultaneous experimentation more common in agriculture). They considered two strategies: augmenting the best main effects plan for k factors run by run and deleting runs one by one from the complete design. In both cases, the run added or deleted is done so on the basis of D-optimality.

Fries and Hunter (1980) proposed the concept of minimum aberration to compare 2^{k-p} designs of equal resolution. This concept generalizes the notion of design resolution, which characterizes factorial designs by stating what high-order interactions need all be negligible in order to assure that all main effects and low-order interactions can be estimated. (For example, a design is said to have resolution III if all the main effects can be estimated, provided that all of the interaction terms are negligible.) Fries and Hunter defined the aberration of a design as the number of words of minimal length in the defining relation for the design and gave examples in which this could be used to compare designs of equal resolution.

Srivastava and Gupta (1979) considered the use of resolution III 2^{k-p} designs when some of the interaction terms are present. In particular, they proposed designs that allow for the detection and estimation of an interaction term, assuming that no more than one interaction is nonnegligible.

Galil and Kiefer (1980a, 1982) thoroughly studied the problem of D-optimal design for weighing experiments and gave extensive tables of the known D-optimal designs. The objective of weighing experiments is to determine the individual weights of k objects in n weighings. For each weighing, each object must be placed in the right pan of the scale, in the

left pan, or not weighed. By identifying the right and left pans of the scale with the two levels of a factor, the weighing design model can be seen as a general paradigm for two-level factorial experiments in n runs; in particular, the 2^{k-p} experiments are a special subset of the general weighing design problem. Galil and Kiefer creatively combined theoretical calculations with computer search to prove the D-optimality of some previously suggested designs and to derive new D-optimal weighing designs. Special attention was given to the most difficult case: $n \equiv 3 \pmod{4}$. Cheng (1980b) showed that certain weighing designs, including fractional factorials, are optimal with respect to a very general class of criteria.

Group screening designs are useful when a large number of factors must be considered and it is desired to find the most important factors with a minimum of experimental runs. Mauro and Smith (1982) investigated the efficiency of two-stage group screening designs in which potentially similar factors are treated as a single factor and varied in unison during a first stage experiment; a second experiment then studies the significant factor groups in detail. Mauro and Smith found that these designs performed quite well, both in terms of identifying significant effects and minimizing the number of runs, even when the initial grouping is based on little prior knowledge.

Several authors have described low-resolution plans for other factorial designs. Anderson and Thomas (1979) gave resolution IV designs for s^k factorials, where s is a power of a prime. The designs require $s(s-1)k$ runs, which is near the theoretical lower bound for an s^k experiment to have resolution IV. Chacko, Dey, and Ramakrishna (1979) derived orthogonal main effect plans for $4^3 2^k$ experiments and showed how they could also be used to construct orthogonal main effect plans for $4^r 3^s 2^k$ experiments when $2 < r + s < 3$. Gupta, Nigam, and Dey (1982) derived orthogonal main effect plans for $t^s k$ factorial experiments.

Cyclic designs have proven to be a useful method to generate blocking schemes for general factorial designs. These designs exploit the theory of cyclic groups and are quite easy to construct. The construction and analysis of cyclic designs for symmetric factorials was described in John and Dean (1975); Dean and John (1975) extended the theory to asymmetrical factorials. The latter article also listed designs for various factor combinations and blocking patterns. John, Wolock, and David (1972) presented an extensive catalog of cyclic designs. John (1981) gave a concise list of efficient cyclic designs.

9. BLOCK DESIGNS

Block designs epitomize one of Fisher's basic concepts of the statistical design of experiments: the importance of setting off experimental runs into small groups (blocks) that are highly homogeneous, in order to increase the precision of the experiment. Classical block designs are intended for experiments with a single factor that has many levels, unlike the factorial experiments described in Section 8, which involve many factors, usually at only two or three levels. When there is only one factor, its levels are usually referred to as treatments (or varieties) and the principal goal of the experiment usually involves comparison of the treatments. The purpose of the blocking scheme, then, is to increase the precision of comparisons among the different treatments.

The classic blocking plans are randomized block designs (for blocking a single factor), Latin Squares and their generalizations (for blocking several factors simultaneously), and incomplete block designs (when the number of treatments exceeds the number of experimental units in each block). Detailed descriptions of these and other blocking schemes are available in many books

on experimental design. In particular, the books by Cochran and Cox (1957) and Kempthorne (1952) are good sources; both books list many designs. John (1971) and Davies (1954) also describe many useful block designs.

Block designs have been the subject of much recent statistical research. In particular, recent work has focused on the application of optimal design theory to block designs. This area is especially conducive to optimal design theory because, for many block designs, a linear statistical model and a precise design region can be clearly stated. Thus the criticisms surrounding the application of optimal design theory to response surface studies (see Section 6) are not serious problems for block designs. The remainder of this section will review some recent results.

One problem that has attracted considerable attention is the design of unbalanced incomplete block designs. The most efficient incomplete block designs for studying v treatments in b blocks of k units each are balanced incomplete block designs, in which each pair of treatments occurs jointly in the same number of blocks. Not all combinations of v , b , and k , however, permit the construction of a balanced design. To aid in finding good incomplete block designs when no balanced design exists, John and Mitchell (1977) introduced the concept of regular graph designs. These are incomplete block designs in which each pair of treatments occurs jointly in either λ_1 or λ_2 blocks, where $\lambda_2 = \lambda_1 + 1$. John and Mitchell showed that these designs are related to a regular graph with v nodes and used graph-theoretic methods to study their properties. They proved that many of the regular graph designs possess optimality properties. Cheng (1978a) showed that regular graph designs are optimal with respect to a large class of optimality criteria, and Cheng (1980a) and Jacroux (1980) gave conditions for the existence of E -optimal regular graph designs. Cheng and Gray (1980) showed

that some special types of regular graph designs are also group divisible. Cheng and Wu (1981) extended the notion of regular graph designs to include experiments in which the treatments are not equally replicated.

Jacroux (1982, 1983) investigated incomplete block designs for which the treatments were not equally replicated and derived some sufficient conditions for such designs to be E-optimal. Results on the E-optimality of some balanced and partially balanced incomplete block designs were also given by Constantine (1981, 1982).

Hall and Jarrett (1981) gave tables of incomplete block designs for experiments with many treatments ($10 < v < 60$) but no more than 5 replicates per treatment and block sizes of at most 10. John (1978) described a new balanced incomplete block design for $v = 18$ treatments in $b = 51$ blocks, with six runs per block and 17 replicates of each treatment. The design is resolvable and can be split into useful partially balanced subdesigns.

Other research work has studied optimality properties of designs that simultaneously block several factors. Kiefer (1975) showed with an elegant proof that generalized Youden designs for simultaneous blocking of two sources of variation are optimal with respect to a large class of optimality criteria. Jacroux (1982) gave E-optimal designs for two-way blocking for experiments with unequally replicated treatments. Cheng (1978b) defined Youden hyperrectangles, which are higher-dimensional generalizations of generalized Youden designs and balanced block designs that allow for blocking many sources of variation, and proved various optimality properties for these designs. Cheng (1979) gave methods for their construction. Cheng (1981) showed that in some cases the optimality properties of generalized Youden designs also hold for a less restrictive class of designs. He called these "pseudo-Youden" designs and gave suggestions on how to construct them.

The blocking schemes described thus far assume that the factors to be blocked in an experiment to compare treatments are categorical variables. Often, however, important concomitant variables are continuous in nature. Harville (1974, 1975) and Cook and Thibodeau (1980) have studied the optimal allocation of experimental units to different treatments when there is covariate information available for each unit at the time of assignment.

Several authors have studied the problem of block designs for experiments where the observations may be subject to a correlated error structure. This problem has attracted attention primarily in agricultural experimentation, where observations from physically adjacent plots may be correlated, but is applicable in a broad range of situations. When such plot-to-plot effects are non-directional (i.e., the errors for two neighboring plots both affect each other), Freeman (1981) recommended the use of quasi-complete Latin squares, which are Latin squares with the property that every unordered pair of elements occurs adjacently twice in rows and twice in columns. Sonneman (1982) studied the case when the plot-to-plot effects are directional (i.e., the error for plot i affects the error for plot $i + 1$, but there is no effect in the other direction), as might occur in a repeated measurement experiment. He proved that complete Latin squares, in which every ordered pair of elements occurs adjacently once in rows and once in columns, are D-optimal. Martin (1982) presented regular and treatment-balanced designs for arranging treatments on a torus when the correlations are assumed to follow a second-order stationary lattice process.

Kiefer and Wynn (1981) proposed a two-stage design strategy for block designs with correlated error structures: first, limit consideration to a class of designs known to be efficient in the absence of correlation (such as balanced incomplete block designs); then, choose a design from within that

The blocking schemes described thus far assume that the factors to be blocked in an experiment to compare treatments are categorical variables. Often, however, important concomitant variables are continuous in nature. Harville (1974, 1975) and Cook and Thibodeau (1980) have studied the optimal allocation of experimental units to different treatments when there is covariate information available for each unit at the time of assignment.

Several authors have studied the problem of block designs for experiments where the observations may be subject to a correlated error structure. This problem has attracted attention primarily in agricultural experimentation, where observations from physically adjacent plots may be correlated, but is applicable in a broad range of situations. When such plot-to-plot effects are non-directional (i.e., the errors for two neighboring plots both affect each other), Freeman (1981) recommended the use of quasi-complete Latin squares, which are Latin squares with the property that every unordered pair of elements occurs adjacently twice in rows and twice in columns. Sonneman (1982) studied the case when the plot-to-plot effects are directional (i.e., the error for plot i affects the error for plot $i + 1$, but there is no effect in the other direction), as might occur in a repeated measurement experiment. He proved that complete Latin squares, in which every ordered pair of elements occurs adjacently once in rows and once in columns, are D-optimal. Martin (1982) presented regular and treatment-balanced designs for arranging treatments on a torus when the correlations are assumed to follow a second-order stationary lattice process.

Kiefer and Wynn (1981) proposed a two-stage design strategy for block designs with correlated error structures: first, limit consideration to a class of designs known to be efficient in the absence of correlation (such as balanced incomplete block designs); then, choose a design from within that

class that offers some protection against possible correlation. They considered a "nearest neighbor" correlation structure and defined the class of equineighborhood designs, which involve restrictions on the number of times pairs of treatments can be adjacent to one another. Cheng (1983) presented methods for constructing such designs.

Bechhofer and Tamhane (1981) studied designs for experiments to compare $v - 1$ test treatments with a control treatment. They introduced the concept of balanced treatment incomplete block (BTIB) designs, which are symmetric with respect to the test treatments and between each test treatment and the control. However, the control may be replicated more often than the test treatments. Bechhofer and Tamhane (1983) further studied these designs and gave tables of optimal allocations of runs among v treatments in order to make one- or two-sided confidence statements. Majumdar and Notz (1983) derived designs for this problem with respect to a variety of optimality criteria. Most of their designs belonged to the class of BTIB designs defined by Bechhofer and Tamhane. Constantine (1983) showed that a simple way to generate a design which minimizes the average variance of the treatment-control comparisons is to reinforce a balanced incomplete block design (for the $v - 1$ test treatments) by adding the control treatment to each block. Again, this is a BTIB design.

10. NONLINEAR MODELS

Nonlinear models play an important role in describing physical, chemical, and engineering systems. By nonlinear models, we refer to situations in which the response Y_i from the i th experimental run is described by the model:

$$Y_i = \eta(x_i; \theta) + \epsilon_i, \quad (10.1)$$

where the response function η is a nonlinear function of the parameter

vector θ . Nonlinear models typically arise when the researcher has in mind a theory that describes the effect of the predictors x_i on the observed response y_i . Even if the theory is incomplete, the nonlinear model may be more useful than a competing empirical model as a first approximation because it more effectively captures the main features of the data. It is often possible to obtain more accurate and more parsimonious models by exploiting the researcher's scientific knowledge to suggest a nonlinear model.

Special design problems arise for nonlinear models because the best design depends, in general, on the unknown parameter values. Investigators are thus in the rather paradoxical position of having to know at the design stage the very quantities that they are conducting the experiment to estimate! Two reviews of work on nonlinear models, including experimental design, are Cochran (1973) and Bates and Hunter (1984).

Fisher (1922) was perhaps the first statistician to study experimental design for a nonlinear model. He considered the problem of designing experiments for the estimation of the density of small organisms in a liquid by means of a series of dilutions. Box and Lucas (1959), in a pioneering paper on experimental design for nonlinear models, showed how the D-optimality criterion could be applied by working with a linearized approximation to the nonlinear model and using the experimenter's initial guesses as to the likely values of the parameters. Box and Hunter (1965a, 1965b) advocated a sequential strategy, in which the parameter estimates are updated after each trial and the next design point is then chosen with the aid of the improved estimates. Hill (1980) showed that if a nonlinear model is linear in some of the parameters, then the D-optimal design does not depend on the value of the linear parameters.

In designing experiments to discriminate among several conjectured models, which was discussed in Section 5, special attention has been paid to the case of nonlinear models. This problem typically arises when there are competing theories to explain the effect of the predictors on the response, each of which implies a different model function η . Experiments are then desired that can discriminate among the models, and so suggest which of the proposed theories seems to be the most valid. See Hunter and Reiner (1965), Box and Hill (1967), Atkinson and Cox (1974), Atkinson and Fedorov (1975a,b), and Atkinson (1981).

Other research work has considered experimental design for particular types of nonlinear models. For example, Currie (1982) compared different designs for estimating the parameters of the Michaelis-Menten equation, which is often used to model enzyme kinetics.

A number of researchers have studied the design of efficient experiments for quantal response data, in which the probability of observing a response is assumed to be a function (typically nonlinear) of some underlying variables, such as dose or stress. A common goal of quantal response experiments is to estimate a stress at which the probability of observing a response obtains a pre-specified level, such as .5. Robbins and Monro (1951) proposed a sequential design scheme (known as stochastic approximation) for this problem, in which the stress for each experimental run is determined by the stress and the outcome of the previous run. Much subsequent work has continued to exploit a sequential approach (see, for example, Wetherill 1963, Tsutakawa 1972, Chernoff 1975, Owen 1975, and Anbar 1978). Other authors have developed non-sequential design schemes for quantal response experiments. Meeker and Hahn (1977) proposed experimental designs to estimate the probability of response at a specified stress, when it is assumed that the probability at

that stress is close to zero or one, and that the probability of response can be accurately represented by a logistic regression model. Abdelbasit and Plackett (1983) also considered logistic regression models and derived designs that maximize information on the parameters in the model. Maxim, Hendrickson, and Cullen (1977) proposed designs for experiments with two stress variables.

11. FUTURE DIRECTIONS

In the preceding sections we have reviewed statistical research on experimental design; in this section we discuss some areas that we think deserve attention in the years ahead. We will begin our discussion with some areas that are natural outgrowths of the recent efforts in experimental design that were discussed in the preceding sections. Then, in individual subsections, we will discuss some other areas that have not been widely explored: designs for sequential experimentation, considering multiple design objectives, planning experiments in the real world, education, and interactive computer programs for designing experiments.

The increasing awareness of the importance of assumptions upon which statistical methods are based has led to much useful research in design robustness (see Section 5). We believe that the problem of designing experiments that will not be overly sensitive to assumptions should be a key concern of statisticians in coming years.

Factorial designs have traditionally been used to study a relatively small number of factors. In many experiments, however, a large number of factors (perhaps 50 or 100) may be initially suspected to be important. The use of highly saturated or even super-saturated factorial designs to study such systems is a problem that should be studied further. It has been reported that in Japan experiments with more than 100 process variables have been successfully performed in industry. Especially influential in Japan have

been Taguchi's ideas on orthogonal arrays (see Taguchi and Wu 1979, Phadke 1982). Another direction worthy of consideration, suggested by Tukey, is the use of designs that are not orthogonal, but in which the correlations of the parameter estimates are quite small. The idea here is that by sacrificing some orthogonality, it may be possible to gain much in terms of the number of factors that can be studied.

The design of experiments for mixture problems is likely to remain a topic of considerable interest. Some of the particular questions that should stimulate more research are designs to combine both mixture and process variables, designs to study only a limited region in the mixture simplex, and computer algorithms for design, especially when additional constraints on the mixture components yield a complicated design region.

The study of experimental design for nonlinear models has lagged far behind the research devoted to experimental design for linear models. One reason for this scarcity of work is the inherent difficulty, discussed in Section 10, that designs generally depend on the unknown parameter values. None the less, nonlinear models are valuable tools for studying processes in the chemical, physical, and engineering sciences; more research on designing experiments for nonlinear models should certainly be undertaken. One interesting question that has not been studied is the design of experiments for nonlinear models that are proposed as tentative empirical approximations. A good design should then allow for estimation of the proposed model, and should also provide a basis for suggesting modifications to the model so that it will more accurately represent the process under study. Another problem that deserves further attention is the link between empirical models and underlying nonlinear mechanisms. This possibility was first noted by Box and Youle (1955), who found that a fitted response surface model for a chemical experiment suggested a theoretical nonlinear model.

11.1 Designs for Sequential Experimentation

Box and Youle (1955) described the iterative nature of experimentation in terms of a cycle that may be repeated many times in the course of an investigation. It consists of four steps: conjecture (experimenter formulates an idea, hypothesis, model, theory), design (experimenter plans the experiment), experiment (experimenter collects the data), and analysis (experimenter extracts useful information from the data). The analysis will frequently cause the experimenter to modify the original conjecture, or even to completely abandon it, in favor of a better conjecture. A new cycle then begins. In the chemical, physical, and engineering sciences, the time to complete a cycle is most often much less than that required in agricultural research, where the basic concepts of experimental design originated. Many new design possibilities can thus be exploited, since the results of previous experiments are available to aid in the planning of future experiments; however, new problems arise on which some research has been done, but more work is definitely in order.

Response surface methodology has always stressed a sequential approach, as is illustrated by the discussion in Box, Hunter, and Hunter (1978, Chapter 15). Even in response surface studies, however, the experimental plan typically proceeds by stages and when stages involve many runs (as is likely if many factors are involved), useful methods might be proposed to further decompose each stage. For example, the first runs of a stage might indicate that another region in the factor space is more interesting than the region currently being explored, that unexpected interactions or other complications are present, that some transformation of the predictor variables is called for, or that unexpected simplifications seem to be possible (e.g., one or more factors are inert, or a simpler model form is appropriate -- perhaps, though

not necessarily, after transformation). In such cases, it would be useful to design the experiment in such a way that initial plans could be modified well short of completion.

Some authors have considered methods to break down experiments into smaller pieces. Box (1982) showed how blocking strategies could be used to construct a sequential scheme for experimentation. Another strategy that might be useful here is to employ the $3/4$ factorial designs introduced by John (1962). Daniel (1973) studied the "one-at-a-time" approach, in which experimental runs are added sequentially from an overall factorial design to create useful designs at each step. A similar idea was proposed by Smith and Schmoyer (1982) (see Sect. 8). A related problem arises when it is difficult or expensive to alter the settings of some factors in an experiment. Draper and Stoneman (1968) and Dickinson (1974) have studied the problem of designing factorial experiments when it is desired to minimize the number of changes in factor settings.

It has been suggested, with some irony, that the best time to design an experiment is after the experiment has been completed because one then has more knowledge of the process under study -- what variables are important, over what ranges, in what metrics, and so on. By designing experiments sequentially, we can, in a sense, approximate this happy (but impossible) situation by "peeking" at the answer and modifying the design accordingly. Such a sequential approach would be optimal in the sense that the planning of each experimental run takes into account all the information available up to the time it is performed. However, using one of the standard design criteria, in which the settings for each run are precisely specified, the investigator would lose the benefits of randomization. In general, the consequences of such a loss are not known and might be a rewarding topic for future research.

11.2 Considering Multiple Design Objectives

Box and Draper (1975) listed 14 different goals which might be important in designing a response surface experiment. Additional goals were listed by Herzberg (1980). Most of the goals in those lists are potentially important in almost any experiment. And the lists are certainly not exhaustive. Experimenters' purposes are complex and often change to reflect new circumstances. Capturing their goals in mathematical terms is an intellectual challenge. Box (1982) stressed the need to design experiments with all important goals in mind, not just one or two. This point is especially important in light of the influence of optimal design, which usually employs a single criterion function. It is good that computer programs that have been developed to search for optimal designs (see Section 4) also compute and output other characteristics of the best designs found, and not just the single criterion by which they search.

Some useful research might be devoted to exploring which goals make similar demands of a design and which goals make contradictory demands. As an example, one way to achieve precise estimation and to allow adequate checks for lack of fit is to increase the sample size; these goals are complementary. Increasing the sample size, however, contradicts the goal of minimizing cost. Another example concerns experimental designs to discriminate among several nonlinear models. As was pointed out by Hill, Hunter, and Wichern (1968), such designs may be quite inefficient for estimating the parameters of the chosen model. They proposed alternative designs that also took parameter estimation into consideration. Further study might suggest effective compromises which allow several goals to be met reasonably well. Some of the work described in Section 5 has attempted to do this, compromising between efficient estimation of an assumed model and the

ability to estimate a more complicated model. A related issue is the problem of designing an experiment that has more than one response (see Draper and Hunter 1966, 1967). An efficient design for one response may not be efficient for some other response. Again, methods are needed which allow for some compromise, so that a design which is reasonably efficient for all the responses can be achieved. Further work in this area would be welcome.

11.3 Planning Experiments in the Real World

Before any formal experimental plan can be laid out, it is essential to state clearly the goals of the experiment and to discuss possible factors that might substantially affect the experimental results and the ability to generalize from them. Although, realistically, there may be infinitely many factors that might affect the results, the experimental design will be able to study only a small subset of them. Thus decisions must be made as to which factors will be systematically varied and over what ranges, which factors held constant, and which factors that are not subject to control should be observed for possible use as covariates in the analysis. (See Hahn 1982a for a useful discussion.) The likely effect of the factors on the experimental results should also be considered. Sometimes current knowledge of the basic mechanism of the system being studied may suggest a useful nonlinear model and the experiment should then be designed with this model in mind. It is usually hoped that the remaining factors, whose effects will all enter into the "error" term in (5.1), are unimportant. Furthermore, it is hoped that if any of these presumably unimportant factors do have large effects, randomization will succeed in neutralizing them. Sometimes the impact of such lurking variables becomes evident only when further experimentation is unable to replicate the original results and a search for additional important factors

is initiated. The possible presence of influential lurking variables helps explain the importance of replicating scientific results by more than one experimenter.

Every good experimental program should consider the issues mentioned in the preceding paragraph. They are especially important, however, for statisticians who aid in planning experiments, since statisticians will often lack intimate knowledge of the subject area in which the experiment is being conducted. Consulting statisticians have found, as Fisher did, that asking questions to clarify these issues and to learn about the experiment is often valuable, not just for their own enlightenment, but also to force experimenters to explain and justify their ideas. As Cochran and Cox (1957) observed: "The statistician who expects that his contribution to the planning will involve some technical matter in statistical theory finds repeatedly that he makes a much more valuable contribution simply by getting the investigator to explain clearly why he is doing the experiment, to justify the experimental treatments whose effects he proposes to compare, and to defend his claim that the completed experiment will enable its objectives to be realized" (p. 10). An illustrative example is the story in Hunter (1981a) of a successful experimental planning session in which the statistician did no more than to ask the two principal investigators to explain the goals of the experiment. The investigators were surprised to discover that each had a different understanding of the goals, but after 45 minutes of vigorous debate, they had established a clear consensus.

How can a statistician learn about the goals of an experiment? What are the important questions to ask at the initial planning phase of an investigation? How can a statistician elicit information which he may regard as crucial to a good design, but that the experimenter regards as marginal?

Joiner and Pollack (1982, p. 334) listed a number of issues that they have repeatedly found to be important. The importance of good consulting skills and the benefits to be derived from working in this area are underrated by statisticians. For information on statistical consulting, see Boen and Zahn (1982), McCulloch et. al. (1982), Joiner (1982), Zahn and Isenberg (1983), and the references listed therein. One important way for statisticians to help themselves is to learn more about the subject matter field(s) in which they consult. They should continually ask questions about the theory underlying an experiment. A deeper understanding of the basic mechanisms that govern the process being studied can often suggest more efficient ways to design and analyze experiments. We believe that the role of the statistician as a planner of experiments is deserving of special consideration. As one suggestion, statisticians who have designed many experiments might consider sharing some of the things they do that seem to be most helpful to their clients, including techniques they use to ensure that they have a clear understanding of the nature of the experiment.

One of the strengths of statistical experimental design is the ability to view experimentation in terms of abstract mathematical models. This abstract view has allowed statisticians to recognize common ground in experiments that otherwise appear to be quite different and has facilitated the invention of many designs that are useful across a broad range of subject areas. In many practical applications, however, idealized, abstract experimental plans must be tempered by the reality of the particular experimental setting at hand. (See, for example, the discussion in Cox 1958, Chapter 9). In particular, designs that have been derived using mathematical criteria should be used as a guideline, not followed slavishly. Consulting statisticians have often found that a visit to the laboratory, plant, or field where the experiment will

actually be carried out is an invaluable aid in proposing a design. An experimental design must be tailored to fit the experiment and not vice versa.

An experimental design that looks great on paper is of little use if it is not followed. Sometimes, in the middle of an experiment, the investigator discovers that some of the planned runs cannot be made, or that the experiment must be terminated early. Some research on how to proceed when the original experimental plan cannot be carried to completion was reported in Sections 6 and 8, but more is needed.

Statistical consultants who propose experimental designs must cooperate closely with the experimenter, so that the latter clearly understands the design and why it is important. In this regard, it is desirable to stress simplicity in developing new experimental designs, but this is a property that is rarely mentioned. Research indicating what difficulties experimenters encounter in applying frequently advocated designs might be of great use to statisticians who design experiments. One suggestion is that the statistician actually participate in running the experiment, or at least be present during the collection of data, in order to obtain first-hand knowledge of all the unforeseen problems that are encountered. This practice can be especially helpful when there are some statistically important factors that the experimenter regarded as inconsequential, and never mentioned to the statistician.

Some readers may think that the questions raised in this section are too trivial to be the subject of statistical interest. We don't think so. To the contrary, we think that these are the most important questions to address and we encourage more statisticians and scientists to share their experiences with problems they have encountered in planning experimental programs. The article by Hahn (1984) is an excellent example of what we have in mind. His description of six experiments in which he participated as a statistical

consultant illustrates how effectively a well-designed experiment can work and also how ingenuity must often be used to make the design fit the needs of the experimenter. Joiner (1977) described the design and analysis of an experiment with a number of unusual, non-standard problems that had to be solved. See also Hooke (1980), Hunter (1981a,b), and Bishop, Peterson, and Trayser (1982). We think more articles of this nature would benefit all of us.

Discussion of the practical problems encountered in planning real-world experiments, sample surveys, and censuses should be included in the training of every scientist and of every statistics student. Too often these problems are swept aside in an instructor's desire to teach material on the theory, rather than the practice, of statistics.

There are indications that increasing use is being made of statistically designed experiments. In Europe, for example, especially in chemistry, the use of designs is becoming widespread, following the leadership of Phan-Tan-Luu, Carlson, and others (see, for example, Carlson, Lundstet, Phan-Tan-Luu, and Mathieu 1983, Carlson, Nilsson, and Strömqvist 1983, Lazaro, Bouchet, and Jacquier 1977, and Brunel, Itier, Commeyras, Phan-Tan-Luu, and Mathieu 1979).

11.4 Education

The growth of knowledge in experimental design over the last 25 years has been tremendous. Many scientists are now aware that statistically designed experiments can greatly increase the efficiency of their research work, which is to the credit of the many individuals who have worked in the field, as well as to journals such as Technometrics that have adopted as a clear priority the dissemination of statistical advances in the chemical, physical, and engineering sciences.

Far more work in education is absolutely necessary. Many experimenters still have little or no idea of even such basic statistical concepts as blocking, replication, randomization, and factorial design. Every applied statistician has a collection of sad tales in which clients asked them to salvage a poorly-planned experiment with a clever analysis. But the damage done by poor experimental design is irreparable. No amount of analysis can create information where none exists in the first place. By contrast, well-planned experiments often require only simple analyses in order to reach clear, unambiguous conclusions. Yet our impression is that many poorly-planned experiments are performed. Indeed, Mead and Pike (1975), in their review of the use of response surface methodology in the biological sciences, concluded that poorly-planned experiments were more the rule than the exception.

It is important to remember how much we can accomplish as teachers. Statistics should be an essential tool for science and engineering students, but it is often regarded as a subject that is too marginal to include in the curriculum. Perhaps one effective way to convince colleagues in other fields of the value of statistical training would be to increase the emphasis on the design of experiments. Many important ideas in the design of experiments can (and should) be taught in introductory statistics courses for university students.

Elements of experimental design can also be taught to high school, junior high school, and grade school students. The difference between correlation and causation, examples of nonsense correlations, how to set up valid comparative experiments, the weakness of varying one variable at a time, and the efficacy of two-level factorial experiments are useful topics to cover. Students need to be taught ideas and procedures that will help them gather

information to better understand the world around them. For example, Dalia Sredni, a seventh grader in California, won first place in a county science fair by conducting a 2^3 factorial design to study the effects of varying the oven temperature, baking time, and amount of baking soda on the height, consistency, texture, and taste of a cake. Students will welcome the opportunity to plan and conduct experiments of their own choosing. Active learning through experiments can be a refreshing change from the more usual passive learning via reading and listening, allowing students to enjoy the element of surprise and the thrill of discovery.

11.5 Interactive Computer Programs

The development of interactive computer programs to aid in the design of experiments will certainly be a focal point in the years ahead. Researchers currently have at their disposal computer packages which allow them to perform almost any standard method of data analysis. No comparable software in the field of experimental design has achieved such widespread use. Easy-to-use, interactive, computer packages could greatly aid researchers in choosing a good experimental design. More important, just as the availability of computer software has led to a revolution in the kinds of data analyses which researchers now regard as essential professional tools, so will software for experimental design lead to a revolution in researchers' awareness of statistically designed experiments.

Some comments are in order to differentiate between the research that we discussed in Section 4 and the ideas of the preceding paragraph. The research which has been done thus far has been devoted largely to developing numerical algorithms which can generate designs with certain desirable properties (such as D-optimal designs). These algorithms have succeeded in finding many useful

designs; however, they are intended more for use in statistical research than for use by experimenters. What we have in mind here is the development of interactive "expert" software packages that experimenters themselves could use to help them design their experiments, in much the same way they would obtain advice from a statistical consultant.

The development of good computer software for experimental design is not a panacea, any more than the existence of statistical analysis packages has been. The proliferation of sophisticated statistical analyses has included many instances where the use of a statistical technique was ill-advised and led to unjustified conclusions. It is important that the users of experimental design packages have at least some knowledge of the basic statistical principles of experimental design. It is also important that the software be intelligent enough to ask the experimenter many of the same questions that a good statistical consultant would ask and to recommend that a statistician be consulted in special circumstances. Thus the comments of the preceding sections regarding the role of the statistician in planning real-world experiments and as an educator should also be seen as essential companions to the development of computer programs for experimental design.

We believe that the benefits of experimental design software far outweigh the potential hazards. Many experiments could be improved substantially by the use of simple, well-established statistical designs. The existence of good software for experimental design would be a great step toward achieving that goal.

12. SOME RECOMMENDATIONS

We conclude with three sets of recommendations addressed to experimenters and statisticians. The theme that runs through these recommendations is that important advances in the theory and practice of experimental design can be achieved if experimenters and statisticians converse with and learn from one another. If communications were improved, each group could help shape future research in the other's area in significant ways. If such dialogue is to bear fruit, concerted effort will be needed to facilitate visits to each other's "camps" to learn the language, customs, problems, and goals of the other group.

1. Teaching and Learning about Experimental Design. Experimenters: if more of you were aware of the concepts and techniques of statistical experimental design and used them in your work, research efficiency -- the amount of information gained per unit of resources used (money, time, etc.) -- in industry, government, and academia could be improved substantially. You need to learn how statistical methods can be combined with the science and technology you know so that you can use that knowledge more effectively in planning experiments to acquire new information; the feeling among some scientists that statistical methodology is a substitute for that knowledge is an unfortunate misconception. Statisticians do not by any means have all the answers, but they have thought deeply about experimental strategy. Many of you could help yourselves considerably by studying what statisticians have written on the subject of experimental design. Those of you who realize the value of statistically designed experiments could help your colleagues by explaining to them the benefits of such an approach.

Statisticians: whether you are teaching statistics in service courses for students from other departments, in courses for your own students, or

special courses (such as continuing education courses for persons in industry), teach proportionately more design and less analysis than you do now. Units on statistics for high school and grade school students should also emphasize (and perhaps begin with) experimental design. Of the two large areas of statistics, data collection and data analysis, the first is more important. A bad design yields data that contain little information, and no amount of clever analysis can extract much information where little exists. Put more positively, the return on investment in good statistical designs can be quite handsome indeed. Talk to some experimenters who have tried it both ways. They'll give you stories you can tell your classes.

2. Using Experimental Design in Practice. One need only glance through journals such as Science to realize how infrequently statistical principles of experimental design are used in the scientific study of complex systems. Porter and Busch (1978) is an exception that proves this "rule." Research workers often consult statisticians, if at all, only after they have assembled their data and encountered difficulties in analyzing them. In many of these situations, the application of basic statistical principles of experimental design would have generated data that were much more informative, not to mention much easier to analyze. Scientists and engineers could reap great benefits by learning and using these principles.

How useful is statistical experimental design in planning experiments? For those of you who hold managerial positions and would like to gauge the possible benefits, we would like to suggest an experiment. In the next year, divide a suitable group of experimenters in your organization into two subgroups, using randomization and perhaps blocking. Provide one of the subgroups with training in statistical experimental design. The training should emphasize the practical rather than the theoretical aspects of the

subject. At the outset, decide on criteria that will be used to assess the research efficiency of these individuals and how the judging will be done. For example, after the passage of a suitable period of time, the reports written by these experimenters could be judged by a panel of experts. Alternatively, hand out identical assignments to two experimenters or teams of experimenters, one of which uses statistically designed experiments and one of which does not. (If you carry out such an experiment, we would like to know what happened.)

Stories commonly swapped among statistical consultants often end this way: "If they had only come to talk to me before they got themselves into that mess, I could have been so much more helpful." (Incidentally, lawyers, doctors, and counselors of all kinds share this same frustration.) Yet the payoff from good design is frequently so great that it is worth the continued effort needed to convince people to talk to you at an early stage in their work. The general problem is that they need to be educated about statistics. The specific problem of persuasion is sometimes solved by communicating to potential clients "success stories" that feature situations or experimenters they know first-hand. Save such stories and use them.

Consultants: with persistence, with creativity, with good humor, with patience, try to get clients to come to you before they collect their data, so you can give them advice on experimental design and so they can reap the rewards. New consultants: lay some long-term plans to cope with this problem and don't get discouraged. Discouraged consultants: revive your good intentions; talk to consultants who have been able to get clients to come to them early for advice on design and learn from them. Consultants who've been successful in this way: publish some of your tips.

3. Establishing and Extending New Frontiers in Experimental Design.

Experimenters: when you learn about the work that statisticians have done on experimental design, many of you will conclude that it is not useful for the work that you do. In fact, some of you will see an enormous gap separating published statistical work and your own needs. You can help statisticians to do better research if you would communicate your perceptions to them.

Statisticians need feedback, information, and advice from experimenters. What forums can be developed to expedite such communication? At professional meetings -- both those of experimenters and those of statisticians -- special sessions should be organized for discussion of such topics. Space in statistics journals should be made available for communications from experimenters concerning research work that they would like to see statisticians undertake. A model of this type of publication is provided, for example, by Rosenblatt and Spiegelman (1981). In a reciprocal manner, scientific journals should provide space for statisticians to make "guest appearances" as Youden did in a popular series in Industrial and Engineering Chemistry and Hahn has been doing in a similar series in Chemtech.

Scientific investigations have served and must continue to serve as a touchstone for statistical research in experimental design. The crucial insight that experimental design should be a branch of statistics became evident to Fisher because of his close interaction with experimental scientists. As Box (1983) observed of Fisher's work at Rothamsted: "One can clearly see the ideas of randomisation, replication, orthogonal arrangement, blocking, factorial designs, measurement of interactions, confounding, all developing in response to the practical necessities of field experimentation" (p. 5). New ideas in the 1950's on response surface methods and, in more recent times, on mixture designs were similarly stimulated by the needs of

experimenters. Much of the novel and useful work in experimental design has been done by statisticians working with or consulting for experimenters in agriculture and industry.

Statistical research exploits more or less general mathematical abstractions of particular experimental settings. Consequently, it is always helpful for statisticians to study the experimental context that gives meaning and relevance to the mathematics because it may reveal explicit or implicit limitations to the theory that has been developed up to that point. We urge research statisticians interested in breaking new ground in experimental design to consult with and, preferably, work collaboratively with experimenters who are working on worthwhile projects -- as Box did at Imperial Chemical Industries and Fisher did at Rothamsted Experimental Station.

ACKNOWLEDGMENTS

We are grateful to many individuals who have read earlier versions of this paper and have provided us with helpful comments, especially Cuthbert Daniel, Bert Gunter, Gerry Hahn, Andrew F. Siegel, and Ed Sylvestre, whose suggestions led us to think more carefully about some of the issues discussed in the paper.

REFERENCES

- ABDELBASIT, K. M. and PLACKETT, R. L. (1983), "Experimental Design for Binary Data," Journal of the American Statistical Association, 78, 90-98.
- ADDELMAN, S. (1962), "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," Technometrics, 4, 21-46.
- ADDELMAN, S. and KEMPTHORNE, O. (1961), "Some Main Effects Plans and Orthogonal Arrays of Strength Two," Annals of Mathematical Statistics, 32, 1167-1176.
- ANBAR, D. (1978), "A Stochastic Newton-Raphson Method," Journal of Statistical Planning and Inference, 2, 153-163.
- ANDERSON, D. A. and THOMAS, A. M. (1979), "Near Minimal Resolution IV Designs for the s^n Factorial Experiment," Technometrics, 21, 331-336.
- ANDREWS, D. F. and HERZBERG, A. M. (1979), "The Robustness and Optimality of Response Surface Designs," Journal of Statistical Planning and Inference, 3, 249-257.
- ASH, A. and HEDAYAT, A. (1978), "An Introduction to Design Optimality with an Overview of the Literature," Communications in Statistics, A7, 1295-1325.
- ATKINSON, A. C. (1972), "Planning Experiments to Detect Inadequate Regression Models," Biometrika, 59, 275-293.
- ATKINSON, A. C. (1981), "A Comparison of Two Criteria for the Design of Experiments for Discriminating Between Models," Technometrics, 23, 301-305.
- ATKINSON, A. C. (1982), Developments in the Design of Experiments," International Statistical Review, 50, 161-177.
- ATKINSON, A. C. and COX, D. R. (1974), "Planning Experiments for Discriminating Between Models," Journal of the Royal Statistical Society, Ser. B, 36, 321-348 (with discussion).

- ATKINSON, A. C. and FEDOROV, V. V. (1975a), "The Design of Experiments for Discriminating Between Two Rival Models," Biometrika, 62, 57-70.
- ATKINSON, A. C. and FEDOROV, V. V. (1975b), "Optimal Design: Experiments for Discriminating Between Several Models," Biometrika, 62, 289-303.
- ATWOOD, C. L. (1973), "Sequences Converging to D-optimal Designs of Experiments," Annals of Statistics, 1, 342-352.
- BAILEY, R. A., GILCHRIST, F. H. L. and PATTERSON, H. D. (1977), "Identification of Effects and Confounding Patterns in Factorial Designs," Biometrika, 64, 347-354.
- BATES, D. M. and HUNTER, W. G. (1984), "Nonlinear Models," in Encyclopedia of Statistical Sciences, Vol. 5, ed. by S. Kotz and N. L. Johnson, New York: John Wiley & Sons, Inc.
- BECHHOFFER, R. E. and TAMHANE, A. C. (1981), "Incomplete Block Designs for Comparing Treatments with a Control: General Theory," Technometrics, 23, 45-57.
- BECHHOFFER, R. E. and TAMHANE, A. C. (1983), "Design of Experiments for Comparing Treatments with a Control: Tables of Optimal Allocations of Observations," Technometrics, 25, 87-95.
- BICKEL, P. J. and HERZBERG, A. M. (1979), "Robustness of Design Against Autocorrelation in Time I: Asymptotic Theory, Optimality for Location and Linear Regression," Annals of Statistics, 7, 77-95.
- BICKEL, P. J., HERZBERG, A. M. and SCHILLING, M. F. (1981), "Robustness of Design Against Autocorrelation in Time II: Optimality, Theoretical and Numerical Results for the First-order Autoregressive Process," Journal of the American Statistical Association, 76, 870-877.

- BISHOP, T., PETERSEN, B. and TRAYSER, D. (1982), "Another Look at the Statistician's Role in Experimental Planning and Design," The American Statistician, 36, 387-389.
- BOEN, J. R. and ZAHN, D. A. (1982), The Human Side of Statistical Consulting, Belmont, Cal.: Lifetime Learning Publications.
- BOX, G. E. P. (1953), "Non-normality and Tests on Variances," Biometrika, 40, 318-335.
- BOX, G. E. P. (1954), "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," Biometrics, 10, 16-60.
- BOX, G. E. P. (1982), "Choice of Response Surface Design and Alphabetic Optimality," Utilitas Mathematica, 21B, 11-55.
- BOX, G. E. P. (1983), "The Importance of Practice in the Development of Statistics," Mathematics Research Center Technical Report No. 2471. University of Wisconsin-Madison.
- BOX, G. E. P. and DRAPER, N. R. (1959), "A Basis for the Selection of a Response Surface Design," Journal of the American Statistical Association, 54, 622-653.
- BOX, G. E. P. and DRAPER, N. R. (1963), "The Choice of a Second Order Rotatable Design," Biometrika, 50, 335-352.
- BOX, G. E. P. and DRAPER, N. R. (1969), Evolutionary Operation, New York: John Wiley and Sons.
- BOX, G. E. P. and DRAPER, N. R. (1975), "Robust Designs," Biometrika, 62, 347-352.
- BOX, G. E. P. and DRAPER, N. R. (1980), "The Variance Function of the Difference Between Two Estimated Responses," Journal of the Royal Statistical Society, Ser. B, 42, 79-82.

- BOX, G. E. P. and DRAPER, N. R. (1982), "Measures of Lack of Fit for Response Surface Designs and Predictor Variable Transformations," Technometrics, 23, 1-8.
- BOX, G. E. P. and GARDNER, C. J. (1966), "Constrained Designs - Part I: First Order Designs," University of Wisconsin-Madison, Department of Statistics Technical Report No. 89.
- BOX, G. E. P. and HILL, W. J. (1967), "Discrimination Among Mechanistic Models," Technometrics, 9, 57-71.
- BOX, G. E. P. and HUNTER, J. S. (1957), "Multi-factor Designs for Exploring Response Surfaces," Annals of Mathematical Statistics, 28, 195-241.
- BOX, G. E. P. and HUNTER, J. S. (1961a), "The 2^{k-P} Fractional Factorial Designs, Part I," Technometrics, 3, 311-351.
- BOX, G. E. P. and HUNTER, J. S. (1961b), "The 2^{k-P} Fractional Factorial Designs, Part II," Technometrics, 3, 449-458.
- BOX, G. E. P. and HUNTER, W. G. (1962), "A Useful Method of Model Building," Technometrics, 4, 301-318.
- BOX, G. E. P. and HUNTER, W. G. (1965a), "The Experimental Study of Physical Mechanisms," Technometrics, 7, 23-42.
- BOX, G. E. P. and HUNTER, W. G. (1965b), "Sequential Design of Experiments for Nonlinear Models," Proceedings of the IBM Scientific Computing Symposium on Statistics, October 21-23, 1963, 113-137.
- BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978), Statistics for Experimenters, New York: John Wiley and Sons, Inc.
- BOX, G. E. P. and LUCAS, H. L. (1959), "Design of Experiments in Nonlinear Situations," Biometrika, 46, 77-90.

- BOX, G. E. P. and WILSON, K. B. (1951), "On the Experimental Attainment of Optimum Conditions," Journal of the Royal Statistical Society, Ser. B, 13, 1-45 (with discussion).
- BOX, G. E. P. and YOLE, P. V. (1955), "The Exploration and Exploitation of Response Surfaces: An Example of the Link Between the Fitted Surface and the Basic Mechanism of the System," Biometrics, 11, 287-323.
- BOX, J. F. (1978), R. A. Fisher, the Life of a Scientist, New York: John Wiley & Sons, Inc.
- BOX, J. F. (1980), "R. A. Fisher and the Design of Experiments, 1922-1926," The American Statistician, 34, 1-7.
- BRADLEY, R. A. and YEH, C. M. (1980), "Trend-free Block Designs: Theory," Annals of Statistics, 8, 883-893.
- BRUNEL, D., ITIER, J., COMMEYRAS, A., PHAN-TAN-LUU, R. and MATHIEU, D. (1979), "Les Acides Perfluorosulfoniques. II Activation du n-pentane par les Systèmes Superacides du Type $R_FSO_3H-SbF_5$ Recherche des Conditions Optimales dans le Cas des Acides $C_4F_9SO_3H$ et CF_3SO_3H ," Bulletin de la Société Chimique de France, 5-6, II 257-II 263.
- CARLSON, R., LUNDSTEDT, T., PHAN-TAN-LUU, R. and MATHIEU, D. (1983), "On the Necessity of Using Multivariate Methods for Optimization in Synthetic Chemistry: An Instructive Example with the Willgerodt Reaction," Nouveau Journal de Chimie, 7, 315-319.
- CARLSON, R., NILSSON, Å. and STROMQVIST, M. (1983), "Optimum Conditions for Enamine Synthesis by an Improved Titanium Tetrachloride Procedure," ACTA Chemica Scandinavica, B37, 7-13.
- CHACKO, A., DEY, A. and RAMAKRISHNA, G. V. S. (1979), "Orthogonal Main-effect Plans for Asymmetrical Factorials," Technometrics, 21, 269-270.

- CHENG, C. S. (1978a), "Optimality of Certain Asymmetrical Experimental Designs," Annals of Statistics, 6, 1239-1261.
- CHENG, C. S. (1978b), "Optimal Designs for the Elimination of Multi-way Heterogeneity," Annals of Statistics, 6, 1262-1272.
- CHENG, C. S. (1979), "Construction of Youden Hyperrectangles," Journal of Statistical Planning and Inference, 3, 109-118.
- CHENG, C. S. (1980a), "On the E-optimality of Some Block Designs," Journal of the Royal Statistical Society, Ser. B, 42, 199-204.
- CHENG, C. S. (1980b), "Optimality of Some Weighing and 2^n Fractional Factorial Designs," Annals of Statistics, 8, 436-446.
- CHENG, C. S. (1981), "Optimality and Construction of Pseudo-Youden Designs," Annals of Statistics, 9, 201-205.
- CHENG, C. S. (1983), "Construction of Optimal Balance Incomplete Block Designs for Correlated Observations," Annals of Statistics, 11, 240-246.
- CHENG, C. S. and GRAY, L. J. (1980), "A Characterization of Group-divisible Designs and Some Related Results," Annals of Discrete Mathematics, 6, 31-39.
- CHENG, C. S. and WU, C. F. (1981), "Nearly Balanced Incomplete Block Designs," Biometrika, 68, 493-500.
- CHERNOFF, H. (1975), "Approaches in Sequential Design of Experiments," in A Survey of Statistical Design and Linear Models, ed. by J. N. Srivastava, New York: North Holland, 67-90.
- COCHRAN, W. G. (1973), "Experiments for Nonlinear Functions," Journal of the American Statistical Association, 68, 771-781.
- COCHRAN, W. G. and COX, G. M. (1957), Experimental Designs, Second Edition. New York: John Wiley & Sons.

- CONSTANTINE, G. M. (1981), "Some E-optimal Block Designs," Annals of Statistics, 9, 886-892.
- CONSTANTINE, G. M. (1982), "On the E-optimality of PBIB Designs with a Small Number of Blocks," Annals of Statistics, 10, 1027-1031.
- CONSTANTINE, G. M. (1983), "On the Trace Efficiency for Control of Reinforced Incomplete Block Designs," Journal of the Royal Statistical Society, Ser. B, 45, 31-36.
- COOK, R. D. and NACHTSHEIM, C. J. (1980), "A Comparison of Algorithms for Constructing Exact D-optimal Designs," Technometrics, 22, 315-324.
- COOK, R. D. and NACHTSHEIM, C. J. (1982), "Model Robust, Linear-optimal Designs," Technometrics, 24, 49-54.
- COOK, R. D. and THIBODEAU, L. A. (1980), "Marginally Restricted D-Optimal Designs," Journal of the American Statistical Association, 75, 366-371.
- CORNELL, J. A. (1973), "Experiments with Mixtures: A Review," Technometrics, 15, 437-455.
- CORNELL, J. A. (1979), "Experiments with Mixtures: An Update and Bibliography," Technometrics, 21, 95-106.
- CORNELL, J. A. (1981), Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data, New York: John Wiley & Sons.
- CORNELL, J. A. and KHURI, A. I. (1979), "Obtaining Constant Prediction Variance on Concentric Triangles for Ternary Mixture Systems," Technometrics, 21, 147-157.
- COX, D. R. (1958), Planning of Experiments, New York: John Wiley and Sons.
- CURRIE, D. J. (1982), "Estimating Michaelis-Menten Parameters: Bias, Variance and Experimental Design," Biometrics, 38, 907-919.
- DANIEL, C. (1973), "One-at-a-time Plans," Journal of the American Statistical Association, 68, 353-360.

- DANIEL, C. (1976), Applications of Statistics to Industrial Experimentation, New York: John Wiley and Sons.
- DANIEL, C. and WILCOXON, F. (1966), "Factorial 2^{P-Q} Plans Robust Against Linear and Quadratic Trends," Technometrics, 8, 259-278.
- DAVIES, O. L. (ed.) (1954), Design and Analysis of Industrial Experiments, New York: Hafner (Macmillan).
- DEAN, A. M. and JOHN, J. A. (1975), "Single Replicate Factorial Experiments in Generalized Cyclic Designs: II. Asymmetrical Arrangements," Journal of the Royal Statistical Society, Ser. B, 37, 72-76.
- DICKINSON, A. W. (1974), "Some Run Orders Requiring a Minimum Number of Factor Level Changes for the 2^4 and 2^5 Main Effect Plans," Technometrics, 16, 31-37.
- DRAPER, N. R. (1982), "Center Points in Second-order Response Surface Designs," Technometrics, 24, 127-133.
- DRAPER, N. R. and HERZBERG, A. M. (1971), "On Lack of Fit," Technometrics, 13, 231-241.
- DRAPER, N. R. and HERZBERG, A. M. (1979), "Designs to Guard Against Outliers in the Presence or Absence of Model Bias," Canadian Journal of Statistics, 7, 127-135.
- DRAPER, N. R. and HUNTER, W. G. (1966), "Design of Experiments for Parameter Estimation in Multiresponse Situations," Biometrika, 53, 525-533.
- DRAPER, N. R. and HUNTER, W. G. (1967), "The Use of Prior Distributions in the Design of Experiments for Parameter Estimation in Nonlinear Situations: Multiresponse Case," Biometrika, 54, 662-665.
- DRAPER, N. R. and STONEMAN, D. M. (1968), "Factor Changes and Linear Trends in Eight-run Two-level Factorial Designs," Technometrics, 10, 302-311.

- ECCLESTON, J. A. and JONES, B. (1980), "Exchange and Interchange Procedures to Search for Optimal Row-and-column Designs," Journal of the Royal Statistical Society, Ser. B, **42**, 372-376.
- ELFVING, G. (1952), "Optimum Allocation in Linear Regression Theory," Annals of Mathematical Statistics, **23**, 255-262.
- EUBANK, R. L., SMITH, P. L. and SMITH, P. W. (1981), "Uniqueness and Eventual Uniqueness of Optimal Designs in Some Time Series Models," Annals of Statistics, **9**, 486-493.
- EUBANK, R. L., SMITH, P. L. and SMITH, P. W. (1982), "A Note on Optimal and Asymptotically Optimal Designs for Certain Time Series Models," Annals of Statistics, **10**, 1295-1301.
- EVANS, J. W. (1979), "Computer Augmentation of Experimental Designs to Maximize $|X'X|$," Technometrics, **21**, 321-330.
- FEDOROV, V. V. (1972), Theory of Optimal Experiments, trans. and ed. by W. J. Studden and E. M. Klimko, New York: Academic Press.
- FINNEY, D. J. (1945), "Fractional Replication of Factorial Arrangements," Annals of Eugenics, **12**, 291-301.
- FISHER, R. A. (1922), "On the Mathematical Foundations of Theoretical Statistics," Philosophical Transactions of the Royal Society of London, Ser. A, **222**, 309-368.
- FISHER, R. A. (1926), "The Arrangement of Field Experiments," Journal of the Ministry of Agriculture, **33**, 503-513.
- FREEMAN, G. H. (1981), "Further Results on Quasi-complete Latin Squares," Journal of the Royal Statistical Society, Ser. B, **43**, 314-320.
- FRIES, A. and HUNTER, W. G. (1980), "Minimum Aberration 2^{k-p} Designs," Technometrics, **22**, 601-608.

- GALIL, Z. and KIEFER, J. (1977a), "Comparison of Rotatable Designs for Regression on Balls, I (Quadratic)," Journal of Statistical Planning and Inference, 1, 27-40.
- GALIL, Z. and KIEFER, J. (1977b), "Comparison of Design for Quadratic Regression on Cubes," Journal of Statistical Planning and Inference, 1, 121-132.
- GALIL, Z. and KIEFER, J. (1979), "Extrapolation Designs and ϕ_p -optimum Designs for Cubic Regression on the q-ball," Journal of Statistical Planning and Inference, 3, 27-38.
- GALIL, Z. and KIEFER, J. (1980a), "D-optimum Weighing Designs," Annals of Statistics, 8, 1293-1306.
- GALIL, Z. and KIEFER, J. (1980b), "Time- and Space-saving Computer Methods, Related to Mitchell's DETMAX, for Finding D-optimum Designs," Technometrics, 22, 301-313.
- GALIL, Z. and KIEFER, J. (1982), "Construction Methods for D-optimum Weighing Designs when $n \equiv 3 \pmod{4}$," Annals of Statistics, 10, 502-510.
- GUPTA, V. K., NIGAM, A. K. and DEY, A. (1982), "Orthogonal Main-effect Plans for Asymmetrical Factorials," Technometrics, 24, 135-137.
- HADER, R. J. and PARK, S. H. (1978), "Slope-rotatable Central Composite Designs," Technometrics, 20, 413-417.
- HAHN, G. J. (1974-present), "Random Samplings," column in Chemtech.
- HAHN, G. J. (1982a), "Design of Experiments: Industrial and Scientific Applications" in Encyclopedia of Statistical Sciences, Vol. 2, ed. by S. Kotz and N. L. Johnson, New York: John Wiley & Sons, Inc., 349-359.
- HAHN, G. J. (1982b), "Design of Experiments: An Annotated Bibliography," in Encyclopedia of Statistical Sciences, Vol. 2, ed. by S. Kotz and N. L. Johnson, New York: John Wiley & Sons, Inc., 359-366.

- HAHN, G. J. (1984), "Experimental Design in the Complex World," Technometrics, 26,
- HAHN, G. J. and DERSHOWITZ, A. F. (1974), "Evolutionary Operation Today - Some Survey Results and Observations," Applied Statistics, 23, 214-218.
- HALL, W. B. and JARRETT, R. G. (1981), "Nonresolvable Incomplete Block Designs with Few Replicates," Biometrika, 68, 617-627.
- HARE, L. B. (1979), "Designs for Mixture Experiments Involving Process Variables," Technometrics, 21, 159-173.
- HARVILLE, D. (1974), "Nearly Optimal Allocation of Experimental Units Using Observed Covariate Values," Technometrics, 16, 589-599.
- HARVILLE, D. (1975), "Computing Optimum Designs for Covariance Models," in A Survey of Statistical Design and Linear Models, ed. by J. N. Srivastava, New York: North-Holland, 209-228.
- HERZBERG, A. M. (1982), "The Robust Design of Experiments: A Review," SERDICA, 8, 223-228.
- HERZBERG, A. M. and ANDREWS, D. F. (1976), "Some Considerations in the Optimal Design of Experiments in Non-optimal Situations," Journal of the Royal Statistical Society, Ser. B, 38, 284-289.
- HILL, P. D. H. (1980), "D-optimal Designs for Partially Nonlinear Regression Models," Technometrics, 22, 275-276.
- HILL, W. J. and HUNTER, W. G. (1966), "A Review of Response Surface Methodology: A Literature Survey," Technometrics, 8, 571-590.
- HILL, W. J., HUNTER, W. G. and WICHERN, D. W. (1968), "A Joint Design Criterion for the Dual Problem of Model Discrimination and Parameter Estimation," Technometrics, 10, 145-160.
- HOAGLIN, D. C., and WELSCH, R. (1978), "The Hat Matrix in Regression and ANOVA," The American Statistician, 32, 17-22.

- HOOKE, R. (1980), "Getting People to use Statistics Properly," The American Statistician, 34, 39-42.
- HOTELLING, H. (1944), "Some Improvements in Weighing and Other Experimental Techniques," Annals of Mathematical Statistics, 15, 297-306.
- HUBER, P. J. (1975), "Robustness and Designs," in A Survey of Statistical Design and Linear Models, ed. by J. N. Srivastava, New York: North Holland, 287-301.
- HUBER, P. J. (1981), Robust Statistics, New York: John Wiley & Sons.
- HUNTER, W. G. (1981a), "The Practice of Statistics: The Real World is an Idea Whose Time has Come," The American Statistician, 35, 72-76.
- HUNTER, W. G. (1981b), "Six Statistical Tales," The Statistician, 30, 107-117.
- HUNTER, W. G. and MEZAKI, R. (1964), "A Model Building Technique for Chemical Engineering Kinetics," American Institute of Chemical Engineering Journal, 10, 315-322.
- HUNTER, W. G. and REINER, A. M. (1965), "Designs for Discriminating Between Two Rival Models," Technometrics, 7, 307-323.
- JACROUX, M. (1980), "On the E-optimality of Regular Graph Designs," Journal of the Royal Statistical Society, Ser. B, 42, 205-209.
- JACROUX, M. (1982), "Some E-optimal Designs for the One-way and Two-way Elimination of Heterogeneity," Journal of the Royal Statistical Society, Ser. B, 44, 253-261.
- JACROUX, M. (1983), "Some Minimum Variance Block Designs for Estimating Treatment Differences," Journal of the Royal Statistical Society, Ser. B, 45, 70-76.
- JOHN, J. A. (1981), "Efficient Cyclic Designs," Journal of the Royal Statistical Society, Ser. B, 43, 76-80.

- JOHN, J. A. and DEAN, A. M. (1975), "Single Replicate Factorial Experiments in Generalized Cyclic Designs: I. Symmetrical Arrangements," Journal of the Royal Statistical Society, Ser. B, 37, 63-71.
- JOHN, J. A. and MITCHELL, T. J. (1977), "Optimal Incomplete Block Designs," Journal of the Royal Statistical Society, Ser. B, 39, 39-43.
- JOHN, J. A., WOLOCK, F. W. and DAVID, H. A. (1972), "Cyclic Designs," National Bureau of Standards Applied Mathematics Series 62.
- JOHN, P. W. M. (1962), "Three-quarter Replicates of 2^n Designs," Biometrics, 18, 172-184.
- JOHN, P. W. M. (1971), Statistical Design and Analysis of Experiments, New York: Macmillan.
- JOHN, P. W. M. (1978), "A New Balanced Design for Eighteen Varieties," Technometrics, 20, 155-158.
- JOHN, P. W. M. (1979), "Missing Points in 2^n and 2^{n-k} Factorial Designs," Technometrics, 21, 225-228.
- JOHNSON, M. E. and NACHTSHEIM, C. J. (1983), "Some Guidelines for Constructing Exact D-optimal Designs on Convex Design Spaces," Technometrics, 25, 271-277.
- JOINER, B. L. (1977), "Evaluation of Cryogenic Flow Meters: An Example in Non-standard Experimental Design and Analysis," Technometrics, 19, 353-379.
- JOINER, B. L. (1982), "Consulting, Statistical" in Encyclopedia of Statistical Sciences, Vol. 2, ed. by S. Kotz and N. L. Johnson, 147-155.
- JOINER, B. L. and CAMPBELL, C. (1976), "Designing Experiments when Run Order is Important," Technometrics, 18, 249-259.
- JOINER, B. L., and POLLACK, A. K. (1982), "Practicing Statistics or What They Forgot to Say in the Classroom," in Teaching of Statistics and Statistical Consulting, ed. by J. S. Rustagi and D. A. Wolfe, New York: Academic Press, 327-342.

- JONES, B. and ECCLESTON, J. A. (1980), "Exchange and Interchange Procedures to Search for Optimal Designs," Journal of the Royal Statistical Society, Ser. B, **42**, 238-243.
- JONES, E. R. and MITCHELL, T. J. (1978), "Design Criteria for Detecting Model Inadequacy," Biometrika, **65**, 541-551.
- KEMPTHORNE, O. (1952), The Design and Analysis of Experiments, New York: John Wiley and Sons.
- KENNARD, R. W. and STONE, L. A. (1969), "Computer Aided Design of Experiments," Technometrics, **11**, 137-148.
- KIEFER, J. (1958), "On the Nonrandomized Optimality and Randomized Non-optimality of Symmetrical Designs," Annals of Mathematical Statistics, **29**, 675-699.
- KIEFER, J. (1959), "Optimum Experimental Designs," Journal of the Royal Statistical Society, Ser. B, **21**, 272-319, (with discussion).
- KIEFER, J. (1975), "Construction and Optimality of Generalized Youden Designs," in A Survey of Statistical Design and Linear Models, ed. by J. N. Srivastava, New York: North-Holland, 333-353.
- KIEFER, J. (1981), "The Interplay of Optimality and Combinatorics in Experimental Design," The Canadian Journal of Statistics, **9**, 1-10.
- KIEFER, J. and WOLFOWITZ, J. (1959), "Optimum Designs in Regression Problems," Annals of Mathematical Statistics, **30**, 271-294.
- KIEFER, J. and WOLFOWITZ, J. (1960), "The Equivalence of Two Extremum Problems," Canadian Journal of Mathematics, **12**, 363-366.
- KIEFER, J. and WYNN, H. P. (1981), "Optimal Balanced Block and Latin Square Designs for Correlated Observations," Annals of Statistics, **9**, 737-757.
- KUSSMAUL, K. (1969), "Protecting Against Assuming the Wrong Degree in Polynomial Regression," Technometrics, **11**, 677-682.

- LAUTER, E. (1974), "Experimental Design in a Class of Models," Mathematische Operationsforschung und Statistik, 5, 379-398.
- LAZARO, R., BOUCHET, P. and JACQUIER, R. (1977), "Plans d'Expériences. II Optimisation par la Méthode Simplex de la Synthèse d'une Pyrazolone Industrielle," Bulletin de la Société Chimique de France, 11-12, 1171-1174.
- LI, K. C. (1983), "Minimaxity for Randomized Designs: Some General Results," Annals of Statistics, 11, 225-239.
- LINDLEY, D. V. and SMITH, A. F. M. (1972), "Bayes Estimates for the Linear Model," Journal of the Royal Statistical Society, Ser. B, 34, 1-41, (with discussion).
- LUCAS, J. M. (1976), "Which Response Surface Design is Best," Technometrics, 18, 411-417.
- MAJUMDAR, D. and NOTZ, W. I. (1983), "Optimal Incomplete Block Designs for Comparing Treatments with a Control," Annals of Statistics, 11, 258-266.
- MARTIN, R. J. (1982), "Some Aspects of Experimental Design and Analysis when Errors are Correlated," Biometrika, 69, 597-612.
- MAURO, C. A. and SMITH, D. E. (1982), "The Performance of Two-stage Group Screening in Factor Screening Experiments," Technometrics, 24, 325-330.
- MAXIM, L. D., HENDRICKSON, A. D. and CULLEN, D. E. (1977), "Experimental Design for Sensitivity Testing: The Weibull Model," Technometrics, 19, 405-413 (with discussion).
- McCULLOCH, C. E., BOROTO, D. R., MEETER, D., POLLAND, R., and ZAHN, D. A. (1982), "A Holistic Approach to Training Statistical Consultants," Florida State University Statistical Report M641.
- MEYER, P. A. and ANDERSON, V. L. (1966), "Extreme Vertices Design of Mixture Experiments," Technometrics, 8, 447-456, (with discussion).
- PIKE, D. J. (1975), "A Review of Response Surface Methodology from a Biometric Viewpoint," Biometrics, 31, 803-851.

- MEEKER, W. Q. and HAHN, G. J. (1977), "Asymptotically Optimum Over-Stress Tests to Estimate the Survival Probability at a Condition with a Low Expected Failure Probability," Technometrics, 19, 381-404 (with discussion).
- MITCHELL, T. J. (1974a), "An Algorithm for the Construction of D-optimal Experimental Designs," Technometrics, 16, 203-210.
- MITCHELL, T. J. (1974b), "Computer Construction of 'D-optimal' First-order Designs," Technometrics, 16, 211-220.
- MITCHELL, T. J. and BAYNE, C. K. (1978), "D-optimal Fractions of Three-level Factorial Designs," Technometrics, 20, 369-380.
- MORRIS, M. D. and MITCHELL, T. J. (1983), "Two-level Multifactor Designs for Detecting the Presence of Interactions," Technometrics, 25, 345-355.
- MYERS, R. H. (1976), Response Surface Methodology.
- NELDER, J. A., and MEAD, R. (1965), "A Simplex Method for Function Minimization," The Computer Journal, 7, 308-313.
- NIGAM, A. K., GUPTA, S. C. and GUPTA, S. (1983), "A New Algorithm for Extreme Vertices Designs for Linear Mixture Models," Technometrics, 25, 367-371.
- O'HAGAN, A. (1978), "Curve Fitting and Optimal Design for Prediction," Journal of the Royal Statistical Society, Ser. B, 40, 1-41, (with discussion).
- OWEN, R. J. (1975), "A Bayesian Sequential Procedure for Quantal Response in the Context of Adaptive Mental Testing," Journal of the American Statistical Association, 70, 351-356.
- PATTERSON, H. D. (1976), "Generation of Factorial Designs," Journal of the Royal Statistical Society, Ser. B, 38, 175-179.
- PATTERSON, H. D. and FAILEY, R. A. (1978), "Design Keys for Factorial Experiments," Applied Statistics, 27, 335-343.
- PAZMAN, A. (1980), "Some Features of the Optimal Design Theory - A Survey," Mathematische Operationsforschung und Statistik, Ser. Stat., 11, 415-446.

- PEARSON, E. S. (1931), "The Analysis of Variance in Cases of Non-normal Variation," Biometrika, 23, 114-133.
- PESOTCHINSKY, L. (1978), " Φ_p -optimal Second Order Designs for Symmetric Regions," Journal of Statistical Planning and Inference, 2, 173-188.
- PESOTCHINSKY, L. (1982), "Optimal Robust Designs: Linear Regression in R^k ," Annals of Statistics, 10, 511-525.
- PHADKE, M. S. (1982), "Quality Engineering Using Design of Experiments," paper presented at the Joint Statistical Meetings, August, 1982.
- PIEPEL, G. F. (1983), "Defining Consistent Constraint Regions in Mixture Experiments," Technometrics, 25, 97-101.
- PLACKETT, R. L. and BURMAN, J. P. (1946), "The Design of Optimum Multifactorial Experiments," Biometrika, 33, 305-325.
- PORTER, W. P. and BUSCH, R. L. (1978), "Fractional Factorial Analysis of Growth and Weaning Success in *Peromyscus maniculatus*," Science, 202, 907-910.
- POWELL, M. J. D. (1964), "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives," The Computer Journal, 7, 155-162.
- RAKTOE, B. L., HEDAYAT, A. and FEDERER, W. T. (1981), Factorial Designs, New York: John Wiley & Sons.
- ROBBINS, H. and MONRO, S. (1951), "A Stochastic Approximation Method," Annals of Mathematical Statistics, 29, 400-407.
- ROSENBLATT, J. R. and SPIEGELMAN, C. H. (1981), Discussion to "A Bayesian Analysis of the Linear Calibration Problem" by W. G. Hunter and W. F. Lamboy, Technometrics, 23, 329-333.
- SACKS, J. and YLVIKAKER, D. (1966), "Designs for Regression Problems with Correlated Errors," Annals of Mathematical Statistics, 37, 66-89.

- SACKS, J. and YLVIKAKER, D. (1968), "Designs for Regression Problems with Correlated Errors; Many Parameters," Annals of Mathematical Statistics, 39, 49-69.
- SACKS, J. and YLVIKAKER, D. (1978), "Linear Estimation for Approximately Linear Models," Annals of Statistics, 6, 1122-1137.
- SCHEFFE, H. (1958), "Experiments with Mixtures," Journal of the Royal Statistical Society, Ser. B, 20, 344-360.
- SCHEFFE, H. (1963), "The Simplex-centroid Design for Experiments with Mixtures," Journal of the Royal Statistical Society, Ser. B, 25, 235-263 (with discussion).
- SILVEY, S. D. (1980), Optimal Design, New York: Chapman and Hall.
- SMITH, A. F. M. and VERDINELLI, I. (1980), "A Note on Bayes Designs for Inference Using a Hierarchical Linear Model," Biometrika, 67, 613-619.
- SMITH, D. E. and SCHMOYER, D. D. (1982), "First-order 'Interruptible' Designs," Technometrics, 24, 55-58.
- SMITH, K. (1918), "On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance They Give Towards a Proper Choice of the Distribution of Observations," Biometrika, 12, 1-85.
- SNEE, R. D. (1979), "Experimental Designs for Mixture Systems with Multicomponent Constraints," Communications in Statistics - Theory and Methods, A8, 303-326.
- SNEE, R. D. and MARQUARDT, D. W. (1974), "Extreme Vertices Designs for Linear Mixture Models," Technometrics, 16, 399-408.
- SONNEMAN, E. (1982), "D-optimality of Complete Latin Squares," Mathematische Operationsforschung und Statistik, Ser. Stat., 13, 387-394.

- SPENDLEY, W., HEXT, G. R. and HIMSWORTH, F. R. (1962), "Sequential Applications of Simplex Designs in Optimization and EVOP," Technometrics, 4, 441-461.
- SRIVASTAVA, J. N. and GUPTA, B. C. (1979), "Main Effect Plan for 2^m Factorials which Allow Search and Estimation of One Unknown Effect," Journal of Statistical Planning and Inference, 3, 259-265.
- STIGLER, S. M. (1971), "Optimal Experimental Design for Polynomial Regression," Journal of the American Statistical Association, 66, 311-318.
- STUDDEN, W. J. (1982), "Some Robust-type D-optimal Designs in Polynomial Regression," Journal of the American Statistical Association, 77, 916-921.
- St. JOHN, R. C. and DRAPER, N. R. (1975), "D-optimality for Regression Designs: A Review," Technometrics, 17, 15-23.
- TAGUCHI, G. and WU, Y. (1979), Introduction to Off-Line Quality Control, Japan: Central Japan Quality Control Association.
- TSUTAKAWA, R. K. (1972), "Design of Experiment for Bioassay," Journal of the American Statistical Association, 67, 584-590.
- VUCHKOV, I. N., DAMGALIEV, D. L. and YONTCHEV, Ch. A. (1981), "Sequentially Generated Second Order Quasi D-optimal Designs for Experiments with Mixture and Process Variables," Technometrics, 23, 233-238.
- WAHBA, G. (1971), "On the Regression Design Problem of Sacks and Ylvisaker," Annals of Mathematical Statistics, 42, 1035-1053.
- WALD, A. (1943), "On the Efficient Design of Statistical Investigations," Annals of Mathematical Statistics, 14, 143-140.
- WELCH, W. J. (1982), "Branch-and-bound Search for Experimental Designs Based on D Optimality and Other Criteria," Technometrics, 24, 41-48.

- WETHERILL, G. B. (1963), "Sequential Estimation of Quantal Response Curves," Journal of the Royal Statistical Society, Ser. B, 25, 1-48 (with discussion).
- WU, C. F. (1981a), "Iterative Construction of Nearly Balanced Assignments I: Categorical Covariates," Technometrics, 23, 37-44.
- WU, C. F. (1981b), "On the Robustness and Efficiency of Some Randomized Designs," Annals of Statistics, 9, 1168-1177.
- WU, C. F. and WYNN, H. P. (1978), "The Convergence of General Step-length Algorithms for Regular Optimum Design Criteria," Annals of Statistics, 6, 1273-1285.
- WYNN, H. P. (1972), "Results in the Theory and Construction of D-optimum Experimental Designs," Journal of the Royal Statistical Society, Ser. B, 34, 133-147.
- YATES, F. (1964), "Sir Ronald Fisher and the Design of Experiments," Biometrics, 20, 307-321.
- YATES, F. and MATHER, K. (1963), "Ronald Aylmer Fisher," Biographical Memoirs of Fellows of the Royal Society, 9, 91-129.
- YOUNDEN, W. J. (1954-1959), Statistical methods column in Industrial and Engineering Chemistry.
- ZAHN, D. A. and ISENBERG, D. J. (1983), "Nonstatistical Aspects of Statistical Consulting," The American Statistician, 37, 297-302.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|--|--|--|
| 1. REPORT NUMBER #2639 | 2. GOVT ACCESSION NO. AD-A139268 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) Experimental Design: Review and Comment | | 5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) David M. Steinberg and William G. Hunter | | 8. CONTRACT OR GRANT NUMBER(s) MCS-8210950 DAAG29-80-C-0041 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of 610 Walnut Street Madison, Wisconsin 53706 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics & Probability |
| 11. CONTROLLING OFFICE NAME AND ADDRESS See Item 18 below | | 12. REPORT DATE February 1984 |
| | | 13. NUMBER OF PAGES 84 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES U. S. Army Research Office P. O. Box 12211 Research Triangle Park North Carolina 27709 National Science Foundation Washington, DC 20550 | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Experimental Design, Optimal Design, Computer-Aided Design, Robustness, Response Surfaces, Mixtures, Factorial Design, Blocking, Nonlinear Models, Consulting, Teaching | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) We review major developments in the design of experiments, offer our thoughts on important directions for the future, and make specific recommenda- tions for experimenters and statisticians who are students and teachers of experimental design, practitioners of experimental design, and researchers jointly exploring new frontiers. Specific topics covered are optimal design, computer-aided design, robust design, response surface design, mixture design, factorial design, block design, and designs for nonlinear models. | | |

